CREDIT-BASED INSURANCE SCORES:
IMPACTS ON CONSUMERS OF AUTOMOBILE INSURANCE

A Report to Congress by the
Federal Trade Commission

July 2007

FEDERAL TRADE COMMISSION

Deborah Platt Majoras	Chairman
Pamela Jones Harbour	Commissioner
Jon Leibowitz	Commissioner
William E. Kovacic	Commissioner
J. Thomas Rosch	Commissioner

Bureau of Economics

Michael R. Baye	Director
Paul A. Pautler	Deputy Director for Consumer Protection
Jesse B. Leary	Assistant Director, Division of Consumer Protection

Bureau of Consumer Protection

Lydia B. Parnes	Director
Mary Beth Richards	Deputy Director
Peggy Twohig	Associate Director, Division of Financial Practices
Thomas B. Pahl	Assistant Director, Division of Financial Practices

Analysis Team

Matias Barenstein, Economist, Bureau of Economics, Div. of Consumer Protection
Archan Ruparel, Research Analyst, Bureau of Economics, Div. of Consumer Protection
Raymond K. Thompson, Research Analyst, Bureau of Economics, Div. of Consumer Protection

Other Contributors

Erik W. Durbin, Dept. Assistant Director, Bureau of Economics, Div. of Consumer Protection
Christopher R. Kelley, Research Analyst, Bureau of Economics, Div. of Consumer Protection
Kenneth H. Kelly, Economist, Bureau of Economics, Div. of Consumer Protection
Michael J. Pickford, Research Analyst, Bureau of Economics, Div. of Consumer Protection
W. Russell Porter, Economist, Bureau of Economics, Div. of Consumer Protection

TABLE OF CONTENTS i

LIST OF TABLES iii

LIST OF FIGURES iv

I. EXECUTIVE SUMMARY 1

II. INTRODUCTION 5

III. DEVELOPMENT AND USE OF CREDIT-BASED INSURANCE SCORES 7

- A. Background and Historical Experience 7
- B. Development of Credit-Based Insurance Scores 12
- C. Use of Credit-Based Insurance Scores 15
- D. State Restrictions on Scores 17

IV. THE RELATIONSHIP BETWEEN CREDIT HISTORY AND RISK 20

- A. Correlation Between Credit History and Risk 20
  - 1. Prior Research 20
  - 2. Commission Research 23
    - a. FTC Database 23
    - b. Other Data Sources 28
- B. Potential Causal Link between Scores and Risk 30

V. EFFECT OF CREDIT-BASED INSURANCE SCORES ON PRICE AND AVAILABILITY 34

- A. Credit-Based Insurance Scores and Cross-Subsidization 35
  - 1. Possible Impact on Car Ownership 39
  - 2. Possible Impact on Uninsured Driving 40
  - 3. Adverse Selection 42
- B. Other Possible Effects of Credit-Based Insurance Scores 46
- C. Effects on Residual Markets for Automobile Insurance 49

VI. EFFECTS OF SCORES ON PROTECTED CLASSES OF CONSUMERS 50

- A. Credit- Based Insurance Scores and Racial, Ethnic, and Income Groups 51
  - 1. Difference in Scores Across Groups 51
  - 2. Possible Reasons for Differences in Scores Across Groups 56
  - 3. Impact of Differences in Scores on Premiums Paid 58
    - a. Effect on Those for Whom Scores Were Available 58
    - b. Effect on Those for Whom Scores Were Not Available 59
- B. Scores as a Proxy for Race and Ethnicity 61
  - 1. Do Scores Act Solely as a Proxy for Race, Ethnicity, or Income? 62
  - 2. Differences in Average Risk by Race, Ethnicity, and Income 64
  - 3. Controlling for Race, Ethnicity, and Income to Test for a Proxy Effect 67
    - a. Existence of a Proxy Effect 67
    - b. Magnitude of a Proxy Effect 69

VII. ALTERNATE SCORING MODELS 73
  A. The FTC Baseline Model 74
  B. Alternative Scoring Models 78
    1. "Race Neutral" Scoring Models 78
    2. Model Discounting Variables with Large Differences by Race and Ethnicity 80

VIII. CONCLUSION 82

TABLES

FIGURES

APPENDIX A. Text of Section 215 of the FACT ACT

APPENDIX B. Requests for Public Comment

APPENDIX C. The Automobile Policy Database

APPENDIX D. Modeling and Analysis Details

APPENDIX E. The Score Building Procedure

APPENDIX F. Robustness Checks and Limitations of the Analysis

TABLES

TABLE 1. Typical Information Used in Credit-Based Insurance Scoring Models

TABLE 2. Claim Frequency, Claim Severity, and Average Total Amount Paid on Claims

TABLE 3. Median Income and Age, and Gender Make-Up, by Race and Ethnicity

TABLE 4. Change in Predicted Amount Paid on Claims from Using Credit-Based Insurance Scores, by Race and Ethnicity

TABLE 5. Estimated Relative Amount Paid on Claims, by Race, Ethnicity, and Neighborhood Income

TABLE 6. Estimated Relative Amount Paid on Claims, by Score Decile, Race, Ethnicity, and Neighborhood Income

TABLE 7. Change in Predicted Amount Paid on Claims from Using Credit-Based Insurance Scores Without and With Controls for Race, Ethnicity, and Income, by Race and Ethnicity

TABLE 8. Change in Predicted Amount Paid on Claims from Using Other Risk Variables, Without and With Controls for Race, Ethnicity, and Income, by Race and Ethnicity

TABLE 9. Baseline Credit-Based Insurance Scoring Model Developed by the FTC

TABLE 10. Credit-Based Insurance Scoring Model Developed by the FTC by Including Controls for Race, Ethnicity, and Neighborhood Income in the Score-Building Process

TABLE 11. Credit-Based Insurance Scoring Model Developed by the FTC Using a Sample of Only Non-Hispanic White Insurance Customers

TABLE 12. Credit-Based Insurance Scoring Model Developed by the FTC by Discounting Variables with Large Differences Across Racial and Ethnic Groups

FIGURES

FIGURE 1. Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

FIGURE 2. Frequency and Average Size (Severity) of Claims, Relative to Highest Score Decile

FIGURE 3. "CLUE" Claims Data: Average Amount Paid Out on Claims, Relative to Highest Score Decile

FIGURE 4. By Model Year of Car: Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile (Property Damage Liability Coverage)

FIGURE 5. Change in Predicted Amount Paid on Claims from Using Scores

FIGURE 6. The Ratio of Uninsured Motorist Claims to Liability Coverage Claims (1996-2003)

FIGURE 7. Share of Cars Insured through States' "Residual Market" Insurance Programs (1996-2003)

FIGURE 8. Distribution of Scores, by Race and Ethnicity

FIGURE 9. Distribution of Race and Ethnicity, by Score Decile

FIGURE 10. Distribution of Scores, by Neighborhood Income

FIGURE 11. Distribution of Neighborhood Income, by Score Decile

FIGURE 12. Distribution of Scores by Race and Ethnicity, After Controlling for Age, Gender, and Neighborhood Income

FIGURE 13. By Race and Ethnicity: Change in Predicted Amount Paid on Claims from Using Scores, by Race and Ethnicity

FIGURE 14. By Race and Ethnicity: Estimated Average Amount Paid Out on Claims, Relative to Non-Hispanic Whites in Highest Score Decile

FIGURE 15. By Neighborhood Income: Estimated Average Amount Paid Out on Claims, Relative to People in Highest Score Decile in High Income Areas

FIGURE 16. Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile, with and without Controls for Race, Ethnicity, and Neighborhood Income

FIGURE 17. FTC Baseline Model - Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

FIGURE 18. Distribution of FTC Baseline Model Credit-Based Insurance Scores, by Race and Ethnicity

FIGURE 19. FTC Score Models with Controls for Race, Ethnicity, and Neighborhood Income: Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

FIGURE 20. Distribution of FTC Credit-Based Insurance Scores, by Race and Ethnicity

FIGURE 21. An Additional FTC Credit-Based Insurance Scoring Model: The "Discounted Predictiveness" Model Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

FIGURE 22. Distribution of FTC Credit-Based Insurance Scores, by Race and Ethnicity

I. EXECUTIVE SUMMARY

Section 215 of the FACT Act (FACTA)[1] requires the Federal Trade Commission (FTC or the Commission) and the Federal Reserve Board (FRB), in consultation with the Department of Housing and Urban Development, to study whether credit scores and credit-based insurance scores affect the availability and affordability of consumer credit, as well as automobile and homeowners insurance. FACTA also directs the agencies to assess and report on how these scores are calculated and used; their effects on consumers, specifically their impact on certain groups of consumers, such as low-income consumers, racial and ethnic minority consumers, etc.; and whether alternative scoring models could be developed that would predict risk in a manner comparable to current models but have smaller differences in scores between different groups of consumers. The Commission issues this report to address credit-based insurance scores[2] primarily in the context of automobile insurance.[3]

Credit-based insurance scores, like credit scores, are numerical summaries of consumers' credit histories. Credit-based insurance scores typically are calculated using information about past delinquencies or information on the public record (*e.g.*, bankruptcies); debt ratios (*i.e.*, how close a consumer is to his or her credit limit); evidence of seeking new credit (*e.g.*, inquiries and new accounts); the length and age of credit history; and the use of certain types of credit (*e.g.*, automobile loans). Insurance companies do not use credit-based insurance scores to predict payment behavior, such as whether premiums will be paid. Rather, they use scores as a factor when estimating the number or total cost of insurance claims that prospective customers (or customers renewing their policies) are likely to file.

---

[1] 15 U.S.C. § 1681 note (2006). Appendix A contains the complete text of Section 215 of the FACT Act.

[2] The FRB will submit a report addressing issues related to the use of credit scores and consumer credit decisions.

[3] The Commission will conduct an empirical analysis of the effects of credit-based insurance scores on issues relating to homeowners insurance; the FTC anticipates that it will submit a report to Congress describing the results of this analysis in early 2008.

Credit-based insurance scores evolved from traditional credit scores, and insurance companies began to use insurance scores in the mid-1990s. Since that time, their use has grown very rapidly. Today, all major automobile insurance companies use credit-based insurance scores in some capacity. Insurers use these scores to assign consumers to risk pools and to determine the premiums that they pay.

Insurance companies argue that credit-based insurance scores assist them in evaluating insurance risk more accurately, thereby helping them charge individual consumers premiums that conform more closely to the insurance risk they actually pose. Others criticize credit-based insurance scores on the grounds that there is no persuasive reason that a consumer's credit history should help predict insurance risk. Moreover, others contend that the use of these scores results in low-income consumers and members of minority groups paying higher premiums than other consumers.

Pursuant to FACTA, the FTC evaluated: (1) how credit-based insurance scores are developed and used; and, in the context of automobile insurance (2) the relationship between scores and risk; (3) possible causes of this relationship; (4) the effect of scores on the price and availability of insurance; (5) the impact of scores on racial and ethnic minority groups and on low-income groups; and (6) whether alternative scoring models are available that predict risk as well as current models and narrow the differences in scores among racial, ethnic, and other particular groups of consumers. In conducting this evaluation, the Commission considered prior research, nearly 200 comments submitted in response to requests for the public's views, information presented in meetings with a variety of interested parties, and its own original empirical research using a database of automobile insurance policies. Based on a careful and comprehensive consideration of this information, the FTC has reached the following findings and conclusions:

- Insurance companies increasingly are using credit-based insurance scores in deciding whether and at what price to offer coverage to consumers.

- Credit-based insurance scores are effective predictors of risk under automobile policies. They are predictive of the number of claims consumers file and the total cost of those claims. The use of scores is therefore likely to make the price of insurance better match the risk of loss posed by the consumer. Thus, on average, higher-risk consumers will pay higher premiums and lower-risk consumers will pay lower premiums.

- Several alternative explanations for the source of the correlation between credit-based insurance scores and risk have been suggested. At this time, there is not sufficient evidence to judge which of these explanations, if any, is correct.

- Use of credit-based insurance scores may result in benefits for consumers. For example, scores permit insurance companies to evaluate risk with greater accuracy, which may make them more willing to offer insurance to higher-risk consumers for whom they would otherwise not be able to determine an appropriate premium. Scores also may make the process of granting and pricing insurance quicker and cheaper, cost savings that may be passed on to consumers in the form of lower premiums. However, little hard data was submitted or available to quantify the magnitude of these benefits to consumers.

- Credit-based insurance scores are distributed differently among racial and ethnic groups, and this difference is likely to have an effect on the insurance premiums that these groups pay, on average.

    - Non-Hispanic whites and Asians are distributed relatively evenly over the range of scores, while African Americans and Hispanics are substantially overrepresented among consumers with the lowest scores (the scores associated with the highest predicted risk) and substantially underrepresented among those with the highest scores.

    - With the use of scores for consumers whose information was included in the FTC's database, the average predicted risk (as measured by the total cost of claims filed) for African Americans

    - and Hispanics increased by 10% and 4.2%, respectively, while the average predicted risk for non-Hispanic whites and Asians decreased by 1.6% and 4.9%, respectively.

- Credit-based insurance scores appear to have little effect as a "proxy" for membership in racial and ethnic groups in decisions related to insurance.

    - The relationship between scores and claims risk remains strong when controls for race, ethnicity, and neighborhood income are included in statistical models of risk.

    - In models with credit-based insurance scores but without controls for race or ethnicity, African Americans and Hispanics are predicted to have average predicted risk 10% and 4.2% higher, respectively, than if scores were not used. In models with scores and with controls for race, ethnicity, and income, these groups have average predicted risk 8.9% and 3.5% higher, respectively than if scores were not used. The difference between these two predictions for African Americans and Hispanics (1.1% and 0.7%, respectively) is a measure of the effect of scores on these groups that is attributable to scores serving as a statistical proxy for race and ethnicity.

    - Several other variables in the FTC's database (*e.g.*, the time period that a consumer has been a customer of a particular firm) have a proportional proxy effect that is similar in magnitude to the small proxy effect associated with credit-based insurance scores.

    - Tests also showed that scores predict insurance risk within racial and ethnic minority groups (*e.g.*, Hispanics with lower scores have higher estimated risk than Hispanics with higher scores). This within-group effect of scores is inconsistent with the theory that scores are solely a proxy for race and ethnicity.

- After trying a variety of approaches, the FTC was not able to develop an alternative credit-based insurance scoring model that would continue to predict risk effectively, yet decrease the differences in scores on average among racial and ethnic groups. This does not mean that a model could not be constructed that meets both of these objectives. It does strongly suggest, however, that there is no readily available scoring model that would do so.

II. INTRODUCTION

Over the past decade, insurance companies increasingly have used information about credit history in the form of credit-based insurance scores to make decisions whether to offer insurance to consumers, and, if so, at what price. Because of the importance of insurance in the daily lives of consumers, the widespread use of these scores raises questions about their impact on consumers. In particular, some have expressed concerns about the effect of scores on the availability and affordability of insurance to members of certain demographic groups, especially racial and ethnic minorities.

In 2003, Congress enacted the Fair and Accurate Credit Transactions Act (FACTA) to make comprehensive changes to the nation's system of handling consumer credit information. In response to concerns that had been raised about credit-based insurance scores, in Section 215 of FACTA Congress directed certain federal agencies, including the FTC, to conduct a broad and rigorous inquiry into the effects of these scores and submit a report to Congress with findings and conclusions. The report is intended to provide policymakers with critical information to enable them to make informed decisions with regard to credit-based insurance scores.

Section 215 of FACTA sets forth specific requirements for studying the effects of credit-based insurance scores in the context of automobile and homeowners insurance. It directs the agencies to include a description of how these scores are created and used, as well as an assessment of the impact of scores on the availability and affordability of automobile and homeowners insurance products. Section 215 also requires a rigorous and empirically sound statistical analysis of the relationship between scores and membership in racial, ethnic, and other protected classes. The mandated study further must evaluate whether scores act as a proxy for membership in racial, ethnic, and other protected classes. Finally, Section 215 requires an analysis of whether scoring models could be constructed that both are effective predictors of risk and result in narrower differences in scores among racial, ethnic, and other protected classes.

Section 215 of FACTA also specifies the process to be used in conducting the study, and the contents of the report to be submitted. The Act directed the agencies to seek input from federal and state regulators and consumer and civil rights organizations, and members of the public concerning methodology and research design. The Act requires the report to include "findings and conclusions of the Commission, recommendations to address specific areas of concerns addressed in the study, and recommendations for legislative or administrative action that the Commission may determine to be necessary to ensure that . . . credit-based insurance scores are used appropriately and fairly to avoid negative effects."[4]

The Commission has conducted a study addressing credit-based insurance scores in the context of automobile insurance. Pursuant to statutory directive, the FTC published two Federal Register Notices[5] soliciting comments from the public concerning methodology and research design. The Commission supplemented this information with numerous discussions between its staff and representatives of other government agencies, private companies, and community, civil rights, consumer, and housing groups. The public comments and information obtained in meetings with the various interested parties provided essential information that allowed the Commission to complete this report. In addition, feedback from state regulators, industry participants, and the consumer, civil rights, and housing groups had a substantial impact on the methodology and scope of the analysis.

---

[4] 15 U.S.C. § 1681 note (2006).

[5] Public Comment on Data, Studies, or Other Evidence Related to the Effects of Credit Scores and Credit-Based Insurance Scores on the Availability and Affordability of Financial Products, 70 Fed. Reg. 9652 (Feb. 28, 2005); Public Comment on Methodology and Research Design for Conducting a Study of the Effects of Credit Scores and Credit-Based Insurance Scores on Availability and Affordability of Financial Products, 69 Fed. Reg. 34167 (June 18, 2004).

This report discusses the information that the FTC considered, its analysis of that information, and its findings and conclusions. Parts I and II above present an Executive Summary and Introduction, respectively. Part III is an overview of the development and use of credit-based insurance scores, and Part IV discusses the relationship between credit history and risk. Part V addresses the effect of credit-based insurance scores on the price and availability of insurance. Part VI explores the impact of credit-based insurance scores on racial, ethnic, and other groups. Part VII describes the FTC's efforts to develop a model that reduces differences for protected classes of consumers while continuing to effectively predict risk. Part VIII is a brief conclusion.

III. DEVELOPMENT AND USE OF CREDIT-BASED INSURANCE SCORES

A. Background and Historical Experience

Consumers purchase insurance to protect themselves against the risk of suffering losses. They tend to be "risk averse," that is, consumers would prefer the certainty of paying the expected value of a loss to the possibility of bearing the full amount of the loss. For example, assume that a driver faces a 1% risk of being in an automobile accident that would cause him or her to suffer a $10,000 loss, which means that the expected value of his or her loss is $100 (1% of $10,000). If the driver is risk averse, he or she would be willing to pay $100 or more to avoid the possible loss of $10,000.

What makes insurance markets possible is that insurance companies do not simply take on the risk of their customers, they actually reduce risk. This does not mean that they reduce the total losses from car accidents or house fires, for example, but rather that they reduce the uncertainty that individuals face without themselves facing nearly the same amount of uncertainty. This is possible because the average loss on a large number of policies can be predicted much more accurately than the losses of a single driver or homeowner. For instance, while it is extremely difficult to predict who among a group of 100,000 drivers will have an accident, it may be possible to predict the total number of accidents for these 100,000 drivers with a low margin of error.[6] By selling many policies that cover the possible losses for many consumers, an insurance company faces much lower uncertainty as to total losses than would each consumer if they did not purchase insurance.

Insurance companies have a strong economic incentive to try to predict risk as accurately as possible. In a competitive market for insurance in which all firms have access to the same information about risk, competition for customers will force insurance companies to offer the lowest rates that cover the expected cost of each policy sold. If an insurance company is able to predict risk better than its competitors, it can identify consumers who currently are paying more than they should based on the risk they pose, and target these consumers by offering them a slightly lower price. Thus, developing and using better risk prediction methods is an important form of competition among insurance companies.

---

[6] This risk reduction is due to the "law of large numbers." Uncertainty is reduced as long as there is a sufficient degree of independence among the risk that individual consumers face. For example, selling flood insurance to those who live in a single flood plain reduces risks less than selling the policies to those who live in a broader geographic area.

For decades, insurance companies have divided consumers into groups based on common characteristics which correlate with risk of loss. Automobile insurance companies divide consumers into groups based on factors such as age, gender, marital status, place of residence, and driving history, among others. Once insurance companies have separated consumers into groups based on these characteristics, they use the average risk of each of these groups in helping to determine the price to charge members of the group.

Insurance companies report that during the last decade they have begun to use credit-based insurance scores to assist them in separating consumers into groups based on risk. Insurers have long used some credit history information when evaluating insurance applications, for example, considering bankruptcy in connection with offering homeowners insurance. In the early 1980s, insurance companies and others began assessing the utility of using additional information about credit history in assessing risk, leading to a more formal use of such information in a fairly simple manner by the early 1990s.[7]

In the early 1990s, Fair Isaac Corporation (Fair Isaac), drawing on its experience developing credit scores, led the initial research to develop credit-based insurance scores. The company developed the first "modern" credit-based insurance score and made it available to insurance companies in 1993.[8] This score was developed to predict the likelihood of claims being submitted for homeowners policies. Fair Isaac introduced a credit-based insurance score for automobile policies in 1995, and ChoicePoint introduced a competing score at about the same time.[9] These scores were developed to predict the loss ratios – claims paid out divided by premiums received – of automobile policies. Following the introduction of these third-party scores, some insurance companies began developing and using their own proprietary scores.

---

[7] Meeting between FTC staff and State Farm (July 13, 2004); Meeting between FTC staff and MetLife Home and Auto (July 12, 2004); Meeting between FTC staff and Allstate (June 23, 2004).

[8] E-mail from Karlene Bowen, Fair Isaac, to Jesse Leary, Assistant Director, Division of Consumer Protection, Bureau of Economics (Jan. 30, 2006) (on file with FTC).

Since the mid-1990s, the use of credit-based insurance scores has grown dramatically. According to industry sources, some of this growth is attributable to changes in technology and industry practices that have made it easier for companies to develop[10] and use these scores.[11] For example, during the 1990s insurance company actuaries began using advanced statistical techniques that made it easier to control for many predictive variables at the same time.[12] This made it easier for them to develop proprietary scores and perhaps made them more receptive to using third-party scores. Insurers also explained that at this time they began combining more and more data from throughout their companies into integrated databases, and this "data warehousing" made it much easier for actuaries and others to engage in the research needed to develop scores.[13]

More fundamentally, however, insurance companies increasingly used credit-based insurance scores because their experience revealed that they were effective predictors of risk. For example, according to a published case study, in the early 1990s, Progressive entered the lower-risk portion of the automobile insurance market. Progressive used sophisticated risk prediction techniques that it had developed in its other lines of business to identify consumers who other insurers were overcharging relative to the risk they posed. Progressive offered these consumers the same coverage at a lower price, thereby persuading some of them to switch to Progressive.[14] The success of Progressive's strategy provided a powerful incentive for incumbent firms to improve their own risk prediction techniques to compete more effectively.[15] Many of them responded to this incentive by increasing their development and use of credit-based insurance risk scores.[16]

---

[9] *Id.*; E-mail from John Wilson, ChoicePoint, to Jesse Leary, Assistant Director, Division of Consumer Protection, Bureau of Economics (June 13, 2005) (on file with FTC).

[10] Developing scores is a fairly expensive process, requiring significant information technology resources and technical expertise. It also requires a large amount of data on loss experience. Many smaller firms, and even some larger firms, therefore do not develop their own scores. *See, e.g.*, Lamont Boyd, Fair Isaac Corporation, Remarks at the Fair Isaac Consumer Empowerment Forum (Sept. 2006) (noting only six firms use a proprietary scoring model).

[11] Industry participants estimate that of the firms that use credit-based risk scores, one-half (as measured by market share) use a proprietary score and one-half use a score that others developed. Among insurers who use a non-proprietary score, about two-thirds use a ChoicePoint score, and one-third use a Fair Isaac score.

[12] These techniques are known as Generalized Linear Models (GLMs). GLMs make it easier to control for many predictive variables at once, and can be used to develop credit-based scoring models. GLMs play a central role in the analysis presented in this report, and are discussed in more detail in Appendix D.

[13] Meeting between FTC staff and The Hartford (July 14, 2004).

Insurance companies now widely use credit-based insurance scores. Today, the fifteen largest automobile insurers (with a combined market share of 72% in 2005) all utilize these scores.[17] Many smaller automobile insurers also use credit-based insurance scores.[18]

The development and increased use of credit-based insurance scores has been accompanied by concerns and criticisms about the validity of the underlying relationship between scores and risk and the fundamental fairness of using credit history information to make decisions about insurance. According to critics, credit-based insurance scores: 1) unfairly penalize consumers who have suffered from medical or economic crises, or who have made perfectly legitimate financial decisions that are penalized by scoring models; 2) affect consumers in arbitrary ways, because credit history information may contain errors; and, 3) have a negative impact on minority and low-income consumers.[19]

---

[14] *See, e.g.*, F. Frei, *Innovation at Progressive (A): Pay as You Go Insurance*, Harv. Bus. Sch. Case Study 9-602-175 (Apr. 29, 2004).

[15] Incumbent firms had an incentive to use the new risk prediction technology in any case. The vigorous competition of Progressive, however, likely spurred incumbent firms to move more aggressively to use this technology than they otherwise would have.

[16] *See id*.

[17] National Association of Insurance Commissioners, "Auto Insurance Database Report 2003/2004" (2006) (on file with the FTC); FTC staff reviews of websites and discussions with industry representatives. No market share data more recent than 2005 was available.

[18] Fair Isaac Corporation states that it sells credit-based insurance scores to roughly 350 firms. Comment from Fair Isaac Corp. to FTC at 14 (Apr. 25, 2005), [hereinafter Fair Isaac Comment], *available at* http://www.ftc.gov/os/comments/FACTA-implementationscorestudy/514719-00090.pdf.

B. Development of Credit-Based Insurance Scores

According to score developers and insurance companies, credit-based insurance scores are developed in the same manner as credit scores generally. To construct a model, score developers obtain a sample of insurance policies for which losses are known. The period of time during which losses occurred or could have occurred is called the "exposure period." Score developers start with the credit information available about customers at the beginning of the exposure period and the known losses for them during the period. Score developers then use various statistical and other techniques to develop a model that predicts losses based on the credit information that was available at the start of the exposure period. If the relationship between the credit information and loss is sufficiently stable over time, the model can be applied to the credit histories of other consumers to predict the risk of loss they pose.

The details of the credit information used in particular models that produce credit-based insurance scores generally are not available. As emphasized above, insurance companies assert that risk prediction techniques are an important form of competition, so firms generally do not want to reveal the credit-based insurance scoring models they use.[20]

---

[19] *Hearing Before the New York State Assembly Comm. on. Ins.* (Oct. 22, 2003) (statement of Birny Birnbaum, Executive Director, Center for Economic Justice).

Some states require by law that insurance companies make their models public. Insurance companies, however, explained that most insurance companies develop and use different scoring models in these states than they use in other states to minimize the competitive disadvantage elsewhere as a result of such mandated disclosures. An important exception is ChoicePoint, which has made its Attract Auto Scoring and other models available to the public.

Based on the information the agency reviewed, a general picture of what data are used in credit-based insurance scoring model emerges.[21] Table 1 presents examples of the types of information that often are used in models to predict credit-based insurance scores. Firms, however, vary significantly in the particular information they use in their models. For example, some insurance companies consider the type of credit granted, while others do not. Moreover, within a category of information, firms may consider different variables in calculating credit-based insurance scores. For instance, an insurance company may use the age of the oldest account in a credit report or may consider the average age of all accounts in the report.

Insurance companies explained that they use credit-based insurance scoring models to predict the amount they will pay out in claims, *i.e.*, claims risk. Some models simply predict the likelihood that a customer will file a claim. These models are most useful in those situations in which credit information is predictive of claim frequency, but not particularly predictive of the size of claims.[22]

---

[20] *See* Comment from National Association of Mutual Insurance Cos. to FTC at 2 (Apr. 25, 2005) [hereinafter NAMIC Comment], *available at* http://www.ftc.gov/os/comments/FACTA-implementationscorestudy/514719- 00088.pdf.

[21] Although credit-based insurance scoring models are developed to predict insurance claims, instead of credit behavior, many of the same types of information are used. A discussion of the factors that Fair Isaac Corporation uses in calculating its credit scores of consumers ("FICO scores") is available at: http://www.myfico.com/CreditEducation/CreditInquiries.

More commonly, however, models are used to predict the "loss ratio,"[23] which is the amount that an insurance company pays out on claims divided by the amount that the customers pay in premiums. This has the advantage of controlling for the effects of non-credit factors on risk, such as age or driving history, as premiums are determined by those other factors. For any particular customer, the loss ratio usually will be either zero (*i.e.*, no claims paid), or a number greater than one (*i.e.*, claims paid in an amount that exceeds premiums received). In contrast, for a group of customers, the loss ratio typically will be a positive number less than one (*i.e.*, some claims paid but in an amount that is less than total premiums received).[24] If there is a strong relationship between customers with a particular credit-related attribute and historic loss ratios, this information can be used to predict the risk of loss associated with a prospective customer who shares that attribute.[25]

Other models are used to predict "pure premiums." Pure premiums are the total amount that an insurance company pays on claims to consumers, not the amount that customers pay in to the company. To build a credit-based insurance scoring model based on pure premiums, it is necessary to control for other risk variables and this can be done in one of two ways. One approach is to scale each consumer's losses by an index of how risky they appear, based on other non-credit risk factors (*e.g.*, age or driving history). This is analogous to the modeling of loss-ratios, with the non-credit-variable risk index playing the role of the premium, but avoids the complications that arise in loss ratio models if a credit score affected the premiums of the policies in the development database.

---

[22] From a technical perspective, modeling frequency is relatively straight-forward. There are a number of standard multivariate techniques that can be used to estimate either the likelihood of a claim occurring, such as logistic regression, or the number of claims that would be expected during a period of time, such as Poisson regression.

[23] Loss ratios can be modeled in a variety of ways. Because loss ratios of individuals have such an oddly-shaped distribution – many zeros and some positive numbers that extend over a wide range – the modeling is not trivial, but it can be handled by GLMs. Loss ratios can also be modeled by decomposing the ratio and modeling the two components – claims paid and premiums – separately. For example, some ChoicePoint models use this technique. *See* e-mail from John Wilson to Jesse Leary, *supra* note 9.

[24] Indeed, for an insurance company to be profitable, the amount that it pays out in claims must be less than the premiums it receives plus its return on investing those premiums.

[25] MetLife has developed a rules-based system under which credit history information is used to sort potential customers based on their predicted loss ratio. MetLife's "Personal Financial Management" uses combinations of various characteristics in an applicant's credit report to assign the applicant to one of several risk categories without ever calculating a numerical score. This type of system essentially is a sophisticated analog to the simple rules-based approach sometimes used prior to the development of credit-based scores, under which, for example, some companies would not write homeowners policies to applicants with recent bankruptcies.

The other approach involves treating credit history variables just like any other variable in predicting risk. One benefit of this approach is that it allows for certain credit history variables to have different effects on predicted risk for different groups of drivers. For example, the age of a consumer's oldest account might be less predictive for young drivers than older drivers. Other credit characteristics might be very informative about drivers without prior claims or violations, but provide limited insight for drivers with poor driving records. Note that this approach may result in a model that does not produce a numerical score based solely on credit history information.

C. Use of Credit-Based Insurance Scores

All insurance companies who use credit-based insurance scores explained that they do so in making decisions concerning *potential* customers. Insurance companies, however, also indicated that their use of scores in policy renewals for *existing* customers is much more varied and complicated. Some states limit the ability of insurance companies to use scores when customers renew policies. Even where not precluded by state law, some insurance companies decide not to use scores when customers renew policies to avoid damaging their relationship with these customers. Other states mandate that firms must use, or must use if the customer requests,[26] updated credit-based insurance scores to modify premium rates. Even where not mandated by state law, some insurance companies use scores to modify premium rates for existing customers on request. In sum, insurance companies use credit-related insurance scores to assess premiums for potential customers and sometimes in determining premiums for existing customers who are renewing their policies.

Insurance companies report that they use credit-based insurance scores in a variety of ways as part of the process of determining whether to offer insurance to prospective customers, and, if so, at what price. Making these determinations usually consists of two steps, referred to as "underwriting" and "rating." In "underwriting," insurance companies use certain characteristics of a consumer to assign him or her to a pool based on the consumer's apparent risk of loss. The pool into which the consumer is placed sets the base premium rate for a policy, with the riskier pools having higher base premium rates. In "rating," the second step, the insurance company uses other risk characteristics to adjust the base premium rate up or down to determine the actual amount the consumer would be charged.[27]

Some insurance companies said that they use credit characteristics in the underwriting step. For example, a firm might assign a potential customer to a risk pool based on the number of claims an applicant has filed in the past several years and the applicant's credit-based insurance score. Using credit-based insurance scores in underwriting thus may affect the premiums that a potential customer would have to pay to obtain coverage, as the risk pool in which the consumer is placed determines his or her base premium rate.

---

[26] *See, e.g.*, R. I. Ins. Regulation 25 § 11 (although requiring firms to recalculate a consumer's score upon request every two years, firms generally can use a change in score only to lower premium rates), *available at* http://www.dbr.state.ri.us/documents/rules/insurance/InsuranceRegulation25.pdf.

[27] There has recently been some movement towards what can be called "continuous rating," in which the risk for each applicant is evaluated and priced without first being assigned to a risk pool, but the two-step process is still standard.

Other insurance companies report that they use scores in the rating step.[28] A simple way to include scores is to determine a consumer's base premium using non-credit factors, such as age or driving history, and then adjust that rate up or down in light of his or her score. A more complex method of using scores is to include credit as a rating factor when developing the entire rating scheme. Such an approach allows credit characteristics to be used interactively with other rating factors. Because how a credit-based insurance score predicts risk may vary with other rating variables, incorporating credit more fully into the rating step may assist in determining premiums that more accurately reflect risk.[29]

D. State Restrictions on Scores

As of June 2006, forty-eight states have taken some form of legislative or regulatory action addressing the use of consumer credit information in insurance underwriting and rating; Pennsylvania and Vermont are the only states that have not regulated insurance scoring.[30] Most of these laws and regulations are based on the

---

[28] While we are not aware that any insurance companies consider credit-based insurance scores at both the underwriting and rating stage, they could do so.

[29] An approach that is intermediate between having credit as an add-on or treating credit like any other rating factor is to make the size of a credit score discount or mark-up depend on other rating variables. For example, the good-credit discount for young single male drivers could be larger or smaller than the good-credit discount for middle-aged married drivers.

[30] The information in this section pertaining to state legislative and regulatory action addressing insurance scoring is from the National Association of Mutual Insurance Companies' (NAMIC) 2004 survey of state laws governing insurance scoring practices. The report is available at:

(continued)

National Conference of Insurance Legislators' (NCOIL) "Model Act Regarding Use of Credit Information in Personal Insurance," which was released in 2002.[31]

The NCOIL Model Act prohibits insurers from using credit information as the sole basis for increasing rates or denying, canceling, or not renewing an insurance policy. The model also prohibits consumer reporting agencies from providing or selling information to others that was submitted to the agency pursuant to an insurance company's inquiry about a consumer's credit information, credit report, or insurance score. Further, the NCOIL model requires insurers to comply with five conditions: insurance companies must (1) notify an applicant for insurance if credit information will be used in underwriting or rating; (2) notify the applicant in the event of an adverse action based on credit information and explain its reasoning for the adverse action; (3) re-write and re-rate a policyholder whose credit report was corrected; (4) indemnify insurance agents and brokers who obtained credit information or insurance scores according to an insurance company's procedures and according to applicable laws and regulations; and (5) file its scoring models with the applicable state department of insurance.[32] Twenty-seven states have adopted laws or regulations that adopt verbatim the language of the NCOIL model or incorporate restrictions that are very similar in scope and nature to those in the NCOIL model.

---

http://www.namic.org/reports/credithistory/credithistory.asp. The information in NAMIC's survey has been updated to reflect newly enacted legislation and regulation through June 2006. Information on this new legislation and regulation is from NAMIC's annual surveys of new state insurance laws and NAMIC's 2007 state law bulletins. The 2005 survey is available at: http://www.namic.org/reports/2005NewLaws/default.asp, the 2006 survey is available at: http://www.namic.org/reports/2006NewLaws/default.asp, and the 2007 state law bulletins are available at: http://www.namic.org/stateLaws/2007stateLawBulletins.asp.

[31] A copy of the text of the NCOIL model is available at: http://www.assureusa.org/docs/NCOIL.doc.

[32] In 2003, the National Association of Insurance Commissions described the NCOIL model in testimony before the U.S. House of Representative, Committee on Financial Services, Subcommittee on Financial Institutions and Consumer Credit. This testimony is available at: http://www.ins.state.ny.us/speeches/pdf/ty030610.pdf.

In addition, twenty-one states have adopted some of the same types of restrictions included in the NCOIL model. Fifteen states prohibit certain uses of credit history information or ban the use of certain negative credit factors in the calculation of an insurance score. Eight states have adopted dispute resolution measures governing an insurance company's responsibility to re-write and re-rate a policyholder whose credit report was corrected. Seven states require insurance companies to notify consumers that their credit information will be used in underwriting or rating. Twelve states require insurers to notify and explain to consumers any adverse action based on credit information. Seven states further require insurers to file their insurance scoring methodologies.

There are several other types of restrictions that have been placed on the use of scores. Three states (Georgia, Illinois, and Utah) prohibit using credit history information as the sole basis in making underwriting or rating decisions. Oregon prohibits the use of credit history information to cancel or not renew existing customers or increase their rates, and Maryland bans the use of credit history when underwriting or rating existing customers.

Finally, four states either have or had effective bans on the use of credit history information in underwriting or rating automobile insurance. Hawaii by statute specifically bans the use of credit information. California and Massachusetts effectively ban the use of scores through their rate regulation processes. Formerly, New Jersey had an effective ban in place, but the use of credit-based insurance scores is now allowed.

IV. THE RELATIONSHIP BETWEEN CREDIT HISTORY AND RISK

Some prior researchers have studied the existence and nature of the relationship between credit history and insurance risk. To explore this relationship, the Commission conducted an analysis of a database of automobile insurance policies that the agency compiled for this study.[33] A consistent finding of prior research and the FTC's analysis is that credit information, specifically credit-based insurance scores, is predictive of the claims made under automobile policies. However, it is not clear what causes scores to be effective predictors of risk.

A. Correlation between Credit History and Risk

1. Prior Research

As discussed above, risk prediction is an important method of competition among insurance firms. Research that insurance companies have conducted about the relationship between credit history and insurance risk therefore typically is proprietary and non-public. Nevertheless, several studies have been made public during the past decade that show a relationship between credit history and insurance risk.

In 2000, James E. Monaghan, an actuary from MetLife Home and Auto, published a study analyzing the relationship between credit history variables and claims on automobile and homeowners insurance policies.[34] He separately assessed a number of credit history variables, including delinquencies, inquiries, and debt utilization rates. Monaghan found that customers with the worst values for these variables posed a greater risk (as measured by loss ratios) than customers with the best values - often roughly 50% more for automobile policies and over 90% more for homeowners policies.[35] He found the same pattern of increased risks when he conducted his analysis controlling for other non-credit risk factors one-by-one.

---

[33] See section IV.A.2 and Appendix C for a description of the database.

[34] James N. Monaghan, *The Impact of Personal Credit History on Loss Performance in Personal Lines, Casualty Actuarial Society Ratemaking Discussion Paper* (2000) (presented at the Winter 2000 CAS forum), *available at* http://www.casact.org/pubs/forum/00wforum/00wf079.pdf

After this research, several insurance industry trade associations hired EPIC Actuaries (EPIC) to construct a database of automobile policies with information from a number of different insurers.[36] EPIC analyzed the link between credit history and risk, and described its results in a report issued in 2003.[37] EPIC reported the relationship between credit scores and different measures of risk. The study showed a strong relationship between credit-based insurance scores and the frequency with which claims were made, as well as between scores and the total dollar amount insurance companies paid on these claims.[38] It also showed: (1) no correlation between scores and the size of liability coverage claims; (2) a weak correlation between scores and the size of collision coverage claims; and (3) a strong correlation between scores and the size of comprehensive coverage claims.

In 2003, researchers at the Bureau of Business Research (BBR) at McCombs School of Business at the University of Texas used data from five automobile insurance companies in Texas to study the relationship between credit-based insurance scores and losses. The BBR researchers found that customers with lower scores were more likely to file claims under their automobile insurance policies than customers with higher insurance scores. In addition, the researchers reported that customers with lower scores filed claims for larger dollar amounts than customers with higher scores.[39] To control for the effects of non-credit risk factors, the BBR researchers used an analysis of loss ratios, and found that loss ratios were higher for customers with lower scores than for customers with higher scores.[40]

---

[35] As discussed in the section on the development of credit scores, the loss ratio can be used to control for the effects of the variables used to determine premiums. However, this relies on the assumption that the premiums accurately reflect the risks associated with those variables.

[36] The automobile policy data that form the core of the database that we used to conduct our analysis for this report are a subset of the data collected for use in the EPIC report. That database is discussed in more detail below, and in Appendix C.

[37] Michael J. Miller and Richard A. Smith, *The Relationship of Credit-Based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity: An Actuarial Study by EPIC Actuaries, LLC* (June 2003) [hereinafter EPIC Study], *available at* http://www.progressive.com/shop/EPIC_CreditScores.pdf.

[38] EPIC also conducted a multivariate analysis that included controls for most non-credit risk variables used to underwrite and rate automobile polices. While the relationship between scores and the total amount paid out on claims was not as large once controls were included, it remained quite strong.

In 2004, the Texas legislature directed the Texas Department of Insurance (TDI) to conduct a study and issue a report addressing the relationship between credit-based insurance scores and risk for automobile and homeowner policies. In reports issued in late 2004 and early 2005,[41] TDI analyzed data from six large insurance firms operating in Texas, using each company's credit scoring model.[42] For automobile policies, it found that scores were negatively correlated with total dollars of claims, *i.e.*, as the scores of customers increased, the total amount that the insurance companies paid out in claims decreased. Insurance companies paid out less on automobile policies for customers with higher scores because they filed fewer claims than customers with lower scores.[43] For homeowners insurance, TDI found similar results. TDI found that scores were negatively correlated with both total dollars of claims and loss ratios, *i.e.*, as the scores of customers increased, the total amount that insurance companies paid out on their policies decreased.

---

[39] Bureau of Business Research, McCombs School of Business, The University of Texas at Austin, "A Statistical Analysis of the Relationship Between Credit History and Insurance Loss" (Mar. 2003). The report does not make clear which particular types of automobile coverage were studied.

[40] *Id.*

[41] Texas Department of Insurance, "Use of Credit Information by Insurers in Texas: The Multivariate Analysis" (Jan. 31, 2005) (supplemental report) [hereinafter 2005 Texas Report]; Texas Department of Insurance, "Use of Credit Information by Insurers in Texas" (Dec. 30, 2004) [hereinafter 2004 Texas Report].

[42] All six insurance companies provided TDI with data on automobile policies, and three of them provided data on homeowners policies.

[43] TDI's findings with regard to automobile policies were consistent regardless of whether it controlled for other risk factors in its analysis.

2. Commission Research

a. FTC Database

The FTC undertook an analysis to determine the relationship between credit history and risk of loss. Five of the firms that provided automobile insurance policy data for the EPIC study described above provided the same information for the Commission's study.[44] This information included policy and driver characteristics, claims, and a ChoicePoint Attract Standard Auto credit-based insurance score for the customer who is named first on the policy. The information submitted to the Commission related to automobile insurance policies in place at any time between July 1, 2000, and June 30, 2001.

The FTC combined this information from insurance companies with data from a number of other sources to create its database. The agency included additional information in the database to broaden the range of credit history variables analyzed; to improve the set of other risk controls in the analysis; to provide an independent measure of claims; and to analyze issues relating to race, ethnicity, income, and national origin.[45] One important feature of the FTC database was that we created weights to make it representative of car owners, by neighborhood income and race and ethnicity, throughout the United States.[46] A more detailed description of the construction and contents of the FTC database is provided in Appendix C.

---

[44] The five firms together represented 27% of the automobile insurance market in 2000. The data were drawn in a way that ensured a nationwide representation of policies. More information about the companies and the database are provided in Appendix C. A discussion of the limitations of the database and of our analysis is presented in Appendix F.

[45] We obtained Fair Isaac credit-based insurance scores for a sub-sample of the people in the database. All of the results presented in the body of the report are for the ChoicePoint Attract score. All of the analysis was also conducted using the Fair Isaac score. The results were qualitatively similar regardless of whether the ChoicePoint or the Fair Isaac score was used. Descriptions of all "robustness checks" and other variations of the analysis are presented in Appendix F.

In assessing the relationship between credit history and risk, the FTC focused its analysis on four major types of coverage included in automobile policies: property damage liability coverage, bodily injury liability coverage, collision coverage, and comprehensive coverage.[47][48] Property damage liability coverage insures the customer against liability for damage he or she causes to the cars and other property of others. Bodily injury liability coverage protects the customer from liability for bodily injuries he or she causes to others. Collision coverage insures the customer against damage to his or her own car from collision or rollover. Comprehensive coverage protects the customer against losses from theft of his or her own car and for damage to the car other than from collision or rollover (*e.g.*, vandalism, fire, hail, etc.).

The FTC first analyzed the simple relationship between credit-based insurance scores and claims for these four coverages. Table 2 shows, for each coverage and for each score decile, the average number of claims per year of coverage (per hundred cars, to show detailed differences across deciles), the average size of claims, and the average total amount paid out on claims per year of coverage (which is the product of the number of claims and the average size of claims).

---

[46] The weighting also makes the data representative by geographic area. See Appendix D for a discussion of the development of the weights.

[47] The FTC database also contains information on two first-party medical coverages, usually referred to as MedPay and personal injury protection, or “PIP.” Claims on these policies are relatively infrequent, and the coverages vary from state to state. For these reasons, we do not focus our analysis on these coverages.

[48] These definitions come from the Insurance Information Institute, and are available in more detail at: http://www.iii.org/individuals/auto/a/basic/.

Figure 1 presents graphs of the relationship between scores and the average total amount paid out on claims. In Figure 1, the horizontal axis shows automobile drivers grouped into ten equal groups ("deciles") based on their credit-based insurance score,[49] with drivers in the decile with lowest scores located at the far left and drivers in the decile with the highest scores at the far right. The vertical axis measures the average dollars paid out on claims per year. This measure of risk is calculated relative to drivers with the highest credit-based insurance scores, which means that the value of the highest-score group (*i.e.*, those in the tenth decile) has been defined as one.

Figure 1 shows that there is a relationship between credit-based insurance scores and risk for all four types of coverage analyzed. Specifically, the downward slopes of the darker (higher) lines in Figure 1 show that as scores increase, the risk of loss consistently decreases. (These lines were produced simply by graphing the average total paid on claims – column (c) – from Table 2, relative to the highest score decile.) They show, for example, that insurance companies paid out nearly twice as much on the property damage liability policies of customers in the group with the lowest scores (*i.e.*, those in the first decile) as they did for the group with the highest scores (*i.e.*, those in the tenth decile). Credit-based insurance scores thus are predictive of the amount that insurance companies pay in claims to consumers.

The FTC then constructed statistical models of insurance claims. These models produce estimates of the relationship between scores and claims, and allow us to control for the effects of other risk variables.

---

[49] Score is measured by deciles because the units of scores are arbitrary, so there is no reason to believe that the relationship between changes in score and changes in risk is constant across the score distribution. For example, going from a score of 600 to 620 may have a different effect on predicted risk than going from 800 to 820.

The lighter (lower) lines in Figure 1 show the relationship between credit-based insurance scores and the amount paid out after controlling for other standard risk factors, such as age and driving history.[50] The slope of each line demonstrates that the relationship between scores and risk persists when controls for other risk variables are included, although the relationship is less strong. Once controls are included, for instance, the amount that insurance companies paid out on property damage liability claims to customers with the lowest credit-based insurance scores was 1.7 times the amount they paid to customers with the highest credit-based insurance scores, down from paying nearly twice as much if no controls are included. Because the relationship is less strong when other variables are included, customers who appear more risky based on non-credit variables are also more likely to have lower credit scores. Nevertheless, even when non-credit variables are included in the analysis, credit-based insurance scores continue to predict the amount that insurance companies are likely to pay out in claims to consumers.

Figure 1 therefore shows that there is a relationship between credit-based insurance scores and the total dollar amount of claims that insurance companies paid. To refine this analysis, the FTC assessed whether customers with the lowest scores were likely to cause insurance companies to pay out more because the customers file more claims, file claims for higher amounts, or both. As shown by the darker (higher) lines in Figure 2, customers with lower scores filed substantially more claims than those with higher scores.[51] For instance, customers with the lowest credit-based insurance scores were about 1.7 times more likely to file a property damage liability claim as customers with the highest credit-based insurance scores. On the other hand, as shown in the lighter (lower) lines in Figure 2, the average size of the claims paid was nearly constant regardless of credit-based insurance score. The one exception is comprehensive coverage, which does show a relationship between claim size and score. The different result for comprehensive coverage may be attributable to a correlation between having a lower score and a higher probability of being a victim of automobile theft, because theft claims are larger than claims resulting from most other events that this type of insurance covers.

---

[50] These other factors are controlled for by estimating a Tweedie GLM model of total dollars of claims using score deciles and all of the other risk factors. Modeling details and the other variables included in the models are discussed in Appendix C. Race, ethnicity, and income are not included at this stage of the analysis.

The underlying claims data presented in Table 2 (which are simple averages without controls for other risk factors) show the same patterns as those in Figures 1 and 2, and provide additional information on the absolute size of claims risk for different coverages and different score deciles. One important point that comes out in Table 2 is the difficulty of predicting the claims of individual customers. While the average number of claims per year in the lowest score decile of collision coverage, for example, was more than twice that in the highest decile, there were still only 12 claims per hundred cars per year of coverage for the lowest score decile. So, the vast majority of customers in even the riskiest decile would not file a claim in a given year. As with other risk variables, credit-based insurance scores are able to separate consumers into groups with different average risk, but cannot predict the claims of individual consumers.

---

[51] The results for the frequency and severity of claims come from models that include controls for other risk variables. Modeling details and the other variables included in the models are discussed in Appendix C.

b. Other Data Sources

In addition to this analysis of the information in the FTC database, the Commission evaluated alternative and independent information to assess the relationship between credit-based insurance scores and risk. ChoicePoint Inc. collects data on claims from most major automobile insurance firms in the United States. The data allow insurance companies to learn whether a potential new customer has filed a claim under a previous policy with another firm, and then use that information in underwriting and rating. ChoicePoint refers to this data set as the Comprehensive Loss Underwriting Exchange ("CLUE").

We obtained the CLUE reports for each person in the FTC database for the period July 1995 – June 2003. This encompasses three time periods: (1) the five years prior to the period of the firm-submitted data; (2) the period of the firm-submitted data (July 2000 – June 2001); and (3) the two-year period following the period of the firm-submitted data. The data on claims prior to the firm-submitted data (*i.e.*, prior to July 2000) were used to construct controls in the risk models that the FTC ran.[52] The CLUE data also give us an alternative and independent source of data on claims to use to measure the relationship between credit-based insurance scores and claims.

Figure 3 shows the average dollars paid out for each decile on policies for each of the four main coverages studied.[53] Each panel includes average claims for three data sources and samples: (1) claims in the data set we received from the firms; (2) claims in CLUE for the year over-lapping with the company data set (July 2000 – June 2001); and, (3) claims in CLUE for the six-month period following the company data set (July 2001 – December 2001).[54]

---

[52] We used three years of prior claims data to construct the risk variables used in the risk models. The use of information on prior claims is an improvement over previously published analyses of credit-based insurance scores, which have not included controls for prior claims filed on policies with consumers' prior insurers.

[53] The results in Figures 1 and 2 are for a stratified sub-sample of the database. The stratification was based on which policies had claims in the company-provided data. The sub-sample is discussed in Appendix C. The results in Figure 3 are for the entire sample of 1.4 million policies. We use the full sample because the stratified sub-sample does not have sufficient information to reliably measure claims in the CLUE data for the six-month period starting July 1, 2001. The results shown on these graphs are not controlled for other

(continued)

These results show a consistent pattern of average total dollars paid out on claims being higher for individuals with lower credit-based insurance scores. The relationship is generally similar across the data sources for the year of overlap, with the exception that it is somewhat weaker for bodily injury liability coverage.[55] For the six months starting July 1, 2001, the results vary for different types of automobile insurance coverage. Comprehensive coverage results look very similar in the two time periods. The overall slope is similar for bodily injury but the relationship is less stable. The relationship becomes much flatter in the later time period for collision coverage, and somewhat flatter for property damage liability. This may be evidence that credit-based insurance scores become less predictive of claims for these coverages as more time passes from when the scores were calculated.

---

non-credit risk variables, because we do not have reliable information about those variables outside of the time period covered by the company data and because CLUE does not contain information at the car level. For the same reasons, we use the sum of the earned car years for each coverage on each policy when analyzing the CLUE data.

[54] We used a six-month period because we were concerned that information on the number of insured vehicles and coverage choices would become less reliable the further in time the data were from the data that the companies provided. We also measured claims for the six-month period starting July 1, 2001, for a sample of drivers limited to those who did not have any claims during the period covered by the company-provided data. This gave results for that time period that were very similar to the results for the full sample for that same time period.

[55] Given the time it can take for the full cost of bodily injury liability claims to be determined, this may affect how claims for bodily injury coverage are reported to the CLUE database.

B. Potential Causal Link between Scores and Risk

Thus, two different data sets, and previously published research, show that credit-based insurance scores are correlated with the total amount that insurance companies pay out on claims under automobile insurance policies.[56] The question that naturally arises is why a customer's credit history makes it more or less likely that he or she will suffer a loss and file an insurance claim. The FTC considered various proposed explanations of such a link and the data available bearing on those explanations. The information available, however, does not allow the agency to draw any broad or definitive explanations why there is a relationship between credit-based insurance scores and risk.

We emphasize that assessing the relationship between credit history and insurance risk necessarily involves addressing the attributes and circumstances *on average* of consumers with particular levels of credit-based insurance scores. Of course, these attributes and circumstances do not necessarily apply to each consumer with a particular level of score. People may have negative information on their credit histories for reasons that would seem to be totally unrelated to insurance risk. The starkest example is when the information is simply incorrect. Consumers also may wind up in financial distress for all sorts of reasons that have no bearing on how risky they are as drivers.[57] In addition, consumers may have credit histories that lead to low scores because of a lack of an extensive credit history. This may reflect societal effects like a lack of mainstream credit offerings where a consumer lives, or a lack of sophistication about mainstream credit markets. Again, it is not apparent that these types of circumstances should lead to higher insurance risk.

---

[56] Section VII of this report contains the results of the FTC's successful efforts to build scoring models that are predictive of risk. The FTC's scoring model predicts risk in the company-provided claims data, and in the CLUE data for an entirely different set of people and a different time period. These results provide additional evidence that credit history information can be used to predict automobile insurance claims.

[57] *Hearing Before the New York State Assembly of Comm. on Ins.* (Oct. 22, 2003) (statement of Birny Birnbaum, Executive Director, Center for Economic Justice).

A strong credit history, however, might indicate that a consumer has taken care in managing his or her financial affairs – avoiding loans that might be difficult to repay, avoiding high balances on credit cards, making sure that bills are not misplaced and are paid on time, etc. A consumer who is prudent in financial matters may also be cautious in other matters related to insurance, such as being more likely to put time, effort, and money into things like car and home maintenance, cautious driving habits, etc. An overall inclination to be prudent may lead a consumer both to have a strong credit history and file fewer insurance claims.

There is ongoing research reflected in the behavioral economics literature that tends to show that people who engage in risky behavior in an area of their lives are often willing to take on more risk in other areas, as well. Researchers have studied attitudes toward risk, as well as behavior, in financial settings and driving, as well as a range of other areas including smoking, occupational choice, and migration.[58] One recent article argues that existing research shows that physiological and psychological factors affect how much risk individuals are willing to take in their financial, driving, and other behavior. Many of the psychological studies surveyed in that article analyze the relationship between psychological factors and risk-taking in a single aspect of life. The authors connect these results between financial behavior and driving from studies on separate groups of people, and posit the theory that credit-based insurance scoring works because scores reflect the psychological makeup of the individual in ways that affect insurance risk.[59]

---

[58] *See, e.g.*, Thomas Dohmen, et al., *Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey* (Sept. 2005), *available at* http://ftp.iza.org/dp1730.pdf.

Others have suggested that credit history provides information about a consumer's circumstances and those circumstances affect the likelihood or size of claims. One example is that a driver with a low credit-based insurance score may be in a distressed financial situation. This may cause stress that makes the consumer a less attentive driver.[60] Being in a distressed financial situation also might give the driver a greater incentive to try to obtain payment under an insurance policy. For example, he or she may be more likely to file a claim for a small amount of damage to an automobile rather than paying for those expenses out of pocket.

Another circumstance that could explain a correlation between credit-based insurance scores and risk of loss under automobile insurance policies is differences in the number of miles driven. The number of miles that a car is driven is directly related to automobile insurance risk, but companies find it difficult to capture information on "miles driven" with a great deal of accuracy. Consumers with lower scores may put more miles on their cars than consumers with higher scores. For example, consumers with lower scores may put more miles on their cars because they have more drivers per car in their household, they share cars with others, etc. If there is a link between credit-based insurance scores and number of miles driven, this could lead to a correlation between credit-based insurance scores and risk.[61]

---

[59] Patrick L. Brackett and Linda L. Golden, *Biological and Psychobehavioral Correlates of Risk Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works*, 74 J. OF RISK AND INS. 23 (2007).

[60] *Id.*

[61] *See, e.g.*, Patrick Butler, *Driver Negligence vs. Odometer Miles: Rival Theories to Explain 12 Predictors of Auto Insurance Claims* (Aug. 9, 2006) (presented at the American Risk & Insurance

(continued)

As discussed above, a circumstance that could explain the relationship between credit-based insurance scores and risk under automobile insurance policies is differences in the resources that consumers put into maintaining their cars. Consumers with lower scores may not be willing or able to spend as much money to maintain their cars. This may, in turn, make the cars more dangerous to operate and lead to more or larger claims. If this were an important part of the explanation for the relationship between scores and risk, one would expect the relationship to be weaker for newer cars, which presumably would not have had the chance to develop maintenance-related safety problems.

The FTC used its database to test this hypothesis. We divided cars in our database into three groups: model years 1992 and older, model years 1993 – 1996, and model years 1997 and later. Using policy information from 2000 to 2001, we estimated the relationship between credit-based insurance scores and property damage liability risk separately for these three groups.[62] Figure 4 shows that credit-based insurance scores are strongly correlated with risk for each group, that is, the slope of the lines reveal that within each of the three model-year categories, consumers with lower scores pose a greater risk of loss than consumers with higher scores.

The relationship between credit-based insurance scores and risk was slightly stronger for the oldest cars. For the oldest cars, consumers with the lowest scores are 1.81 times riskier than consumers with the highest scores. By contrast, for the newest cars, consumers with the lowest scores are 1.68 times riskier than consumers with the highest scores, and for middle-aged cars, consumers with the lowest scores are 1.64 times

---

Association Annual Meeting), *available at* http://www.aria.org/meetings/2006papers/butler.pdf.

[62] We used property damage liability because (unlike collision or comprehensive coverage) the size of claims does not depend on the value of the car covered by the policy. Car values will vary with model year, so using coverages where the size of claims varies with the value of the car would complicate the analysis.

riskier than consumers with the highest scores. Our results are weakly consistent with the hypothesis that some of the relationship is attributable to consumers with lower scores spending less to maintain their vehicles, but also show that difference in maintenance is not the primary cause of the relationship.

In short, many explanations have been offered as to why the characteristics or circumstances of consumers might account for the relationship between scores and risk. Little empirical data testing these possible explanations are available. The FTC tested one possible explanation for the relationship between scores and risk under automobile policies, and the results were weakly consistent with the hypothesis that some of the relationship could be attributable to the lower amount that consumers with lower scores may spend on maintenance. Although this result provides some insight, the information available does not allow the agency to draw any broad or definitive conclusions as to the reason that there is a relationship between scores and risk.

V. EFFECT OF CREDIT-BASED INSURANCE SCORES ON PRICE AND AVAILABILITY

Credit-based insurance scores are predictive of risk for automobile policies. Insurance companies therefore are able to use these scores to underwrite and rate policies in ways that correspond more closely to individual risk, on average. Enhanced accuracy results in decreased premiums for lower-risk consumers and in increased premiums for higher-risk consumers, and reduces the extent to which lower-risk consumers subsidize higher-risk consumers. Enhanced accuracy also may have broader effects in the marketplace. It may make insurance companies willing to offer policies to consumers posing a wider range of risk and it may reduce adverse selection among consumers.

A. Credit-Based Insurance Scores and Cross-Subsidization

Every insurance policy written for a consumer can be thought of as posing a true level of claims risk, that is, the expected cost to the insurance company of claims that the customer will submit. If the firm knew this true level of risk, it could base premiums on this risk. Because of practical limitations on the ability of firms to obtain and process information, they cannot determine the true level of risk that any particular consumer poses.[63] Instead, they must use the information available to them to estimate the expected claims cost for each consumer. Traditionally, insurance companies have divided customers into groups based on their characteristics and calculated expected average losses for the group, after which group members are charged premiums based on these expected losses.

Because the true expected claims costs will vary within any group of customers, some in the group will be paying premiums that are higher and others will be paying premiums that are lower than their own individual true expected claims cost. Those in the group with lower expected claims costs (*i.e.*, the lower-risk customers) subsidize those with the higher expected claims cost (*i.e.*, the higher-risk customers).[64] In the absence of perfect information about individual customer risks, there will always be some consumers in an insured group who subsidize other consumers in the group.[65]

---

[63] Because insurers never have complete information about consumers, their estimates of expected claims costs are, at best, only correct on average; some estimates are over-estimates and others under-estimates. Such a situation is referred to as "imperfect information" about consumer risk.

[64] This is *ex ante* cross-subsidization (or, cross-subsidization "in expectation"). It is a distinct concept from *ex post* cross-subsidization. Inherent in the concept of insurance is *ex post* cross-subsidization, that is, customers who do not experience loss subsidize customers who do.

[65] Note that if information is symmetric between insurers and consumers (*i.e.*, they both have the same imperfect beliefs about expected claims costs), consumers will not know whether they are beneficiaries or contributors to the subsidization. Given that consumers do not know whether they are paying more or less

(continued)

Better risk prediction techniques allow insurance companies to more effectively separate higher-risk consumers from lower-risk consumers. This information assists insurance companies in charging consumers prices that correspond more closely to the true risk they pose, on average. This, in turn, decreases the premiums of lower-risk consumers and increases the premiums of higher-risk consumers, on average. Improved risk prediction techniques therefore reduce the extent to which lower-risk consumers subsidize higher-risk consumers.[66]

Even though improved risk prediction techniques will make firms' estimates of the riskiness of consumers on average more accurate, the predicted risk of some individual consumers may become less accurate. For example, there are some consumers who are very safe drivers but have low credit-based insurance scores. If scores are used, the predicted risk for these specific individuals will become less accurate. This result is unavoidable in any scheme used to make predictions about the risk consumers pose. Therefore, even if risk predictions become more accurate overall as additional predictive information is considered, there will always be some people who are much safer – or much riskier – than they appear.

The FTC analyzed the information in its automobile insurance database to estimate the extent to which the use of credit-based insurance scores (a risk prediction technique) could reduce cross-subsidization. Many of the premiums for policies included in the database were calculated without using scores, [67] and the data do not indicate

---

than their true risk, their decisions will be unaffected by the existence of cross-subsidization.

[66] This is true even for customers of firms that do not adopt the more accurate prediction method, because those firms will wind up with a riskier and more homogenous pool of customers. Because the pool of customers is more homogenous, there will be less cross-subsidization within that group of consumers.

[67] E-mail from Rick Smith, Towers Perrin, to Jesse Leary, Assistant Director, Division of Consumer Protection, Bureau of Economics (Apr. 13, 2005) (on file with FTC).

which policies these were. The FTC database contains information from 2000-2001, shortly after the introduction of scores. As discussed above, scores typically are used in determining the premiums to be charged to *prospective* customers. Customers who renewed their policies during 2000-2001 thus were not likely to have had scores used to determine their premiums. In addition, although by 2000 insurance companies were using scores to determine premiums in many states, their use was not universal. Accordingly, many, and probably most, of the premiums charged to consumers during this period of time were determined without the use of credit-based insurance scores.

Because most of the premiums in the database likely do not reflect the use of credit-based insurance scores, the FTC used risk, measured in expected total dollars of claims, as a substitute for premiums in an analysis of the effects of scores. We believe that this calculation of risk is a reasonable substitute for premiums in this context, because the premiums that an insurance company charges consumers in a competitive marketplace should be roughly proportional to the risk they appear to pose.[68]

The FTC used a three step analysis to evaluate how expected risk changes if insurance companies consider credit-based insurance scores. The first step was to use a model to calculate a predicted dollar risk for each consumer using all risk factors in the database, except score.[69] The second step was to calculate a predicted dollar risk for each consumer using all risk factors plus a score. Both of these steps to calculate predicted dollar risk were conducted separately for property damage liability, bodily injury liability, collision, and comprehensive coverage.[70] The third and final step was to sum the predicted dollar risks for all four types of insurance coverage with and without the use of credit-based insurance scores.[71] This produced two estimates of total risk for each insurance policy in the database: an estimate without using a score, and an estimate using a score.

---

[68] Some industry participants have stated that homeowners and automobile insurance markets are fiercely competitive. *See, e.g.*, Comment from State Farm Ins. Co. to FTC at 3-4 (Apr. 25, 2003) [hereinafter State Farm Comment], *available at* http://www.ftc.gov/os/comments/FACTA-implementationscorestudy/514719-00100.pdf; Comment from the American Ins. Ass'n to FTC at 14 (Apr. 25, 2005) [hereinafter AIA Comment], *available at* http://www.ftc.gov/os/comments/FACTA-implementationscorestudy/514719-00084.

[69] This was done using a Tweedie GLM model. Modeling details are provided in Appendix C. Race, ethnicity, and income were not considered at this stage of the analysis.

The FTC's analysis predicts that the use of credit-based insurance scores redistributes premium costs from consumers with higher scores to those with lower scores.[72] This is a zero-sum calculation: the total increases in premiums predicted if scores are used must be exactly the same as the total decreases in premiums predicted.

Figure 5 shows the results of the FTC's analysis of the effect of credit-based insurance scores on changes in premiums. It shows what share of consumers would be predicted to have changes of different sizes. Figure 5 also reveals that if credit-based insurance scores are used, more consumers (59%) would be predicted to have a decrease in their premiums than an increase (41%).

---

[70] We also conducted this analysis using a single-equation model of all coverages, instead of separate models by coverage. As discussed in Appendix F, this yielded similar results.

[71] This approach uses the actual coverage choices of individuals. That is, we predict claims cost for individuals only for the coverages they had, and measure the change in their total predicted claims for those coverages. This has the advantage of taking into account the real choices people made when purchasing insurance, but the disadvantage of not allowing for the possibility that individuals would change their coverage choices in response to changing premiums. To generate the overall distribution of changes, we weighted consumers by the earned car years on their property damage liability coverage.

[72] We emphasize that this is not a measure of how firms are actually using scores to price consumers. Scores are not used to underwrite or rate all customers, especially existing customers. In addition, firms may not adjust premiums in response to scores as much as our analysis would predict. For these two reasons, this exercise may overstate the redistributive effects of using scores. The fundamental assumption of the analysis, that premiums will be proportional to predicted risk, is likely to be violated in the short-term, especially if existing customers are not fully re-underwritten and re-rated every year. In sum, this approach probably overstates the redistributive effects to date of using scores, but should be a reasonable substitute for the long-term effects of using scores.

The increased premiums for consumers whose premiums would rise are larger than the decreased premiums for those whose premiums would fall. This can be seen in the longer "tail" on the right-hand side of the graph, which shows larger changes in the direction of an increase. The median increase for those with an increase in predicted risk is 16% (*i.e.*, one-half the increases in predicted risk are greater than 16% and one-half are less than 16%), while the median decrease is 13%.

1. Possible Impact on Car Ownership

If using credit-based insurance scores results in consumers paying premiums that are closer to the true risk that they pose, this could result in car owners incurring costs closer to the real costs of owning and operating their cars. Internalizing these costs could affect consumer decisions whether to own cars, thus resulting in more efficient car ownership.

If consumers decide how many cars to own based on the benefits and costs of car ownership, their decisions can be said to be "efficient" in that they will choose to own cars only when the benefits are at least as great as the costs. If consumers pay premiums that are lower than the risk they actually pose**,** they will own more cars than is efficient, because other people are helping to pay for the cost of their driving.[73] And, if consumers pay premiums that are higher than the risk they pose, they will own fewer cars than would be efficient, because they face costs higher than the true total costs of their driving. The use of credit-based insurance scores to charge premiums that more accurately reflect the true cost of driving, on average, thus could lead to a more efficient level of car ownership.

---

[73] This is a classic "negative externality." Negative externalities arise any time consumers or businesses pay costs for a product that are less than the costs to society. Because societal costs are not considered in the decision of consumers or businesses, their decisions will be inefficient for society.

The FTC was not able to determine whether and to what extent credit-based insurance scores have an effect on automobile ownership. We are not aware of information addressing specifically how much of an increase or decrease in the cost of driving will cause consumers to decide whether to own a vehicle.[74] Moreover, even if we were able to determine the effect of insurance scores on car ownership, this study does not assess whether such an outcome would be equitable.

2. Possible Impact on Uninsured Driving

Using credit-based risk scores to determine premiums also could have an effect on the number of drivers who drive without insurance. Although most states have requirements that drivers carry specified minimum amounts of liability insurance, there are still significant numbers of drivers who drive without insurance. Raising premiums of drivers with lower scores could lead to more of them driving without insurance. Lowering premiums of drivers with higher scores could lead to fewer of them driving without insurance. Whether the use of scores on balance leads to more or fewer people driving without insurance depends on which of these two effects is greater.[75]

---

[74] Even though there is published work on the effects of prices on new car sales, *see, e.g.,* Patrick S. McCarthy, *Market Price and Income Elasticity of New Vehicle Demands*, 78 REV. OF ECON. AND STATS. 543 (1996), we are not aware of studies that measure the effect of the cost of insurance on the number of cars that households choose to own.

[75] Although it is not obvious which change would be larger, there is strong intuition that suggests people with higher scores are relatively less likely to be driving without insurance even when scores are not used to determine premiums. This would be true if scores were correlated with wealth or if scores were a measure of caution or responsibility. The value of liability insurance to an individual depends in part on that individual's wealth, because people with very little wealth may be nearly "judgment proof," and therefore face very little effective risk from liability claims. A company that issues a policy, however, is liable up to the policy limits. So, liability insurance may be worth less to a low-wealth driver than it costs,
(continued)

The FTC sought to estimate the impact of credit-based insurance scores on the prevalence of consumers driving without insurance. It is difficult to obtain reliable data concerning the number or share of drivers who drive without insurance, because this conduct is illegal in most states. In an effort to derive such an estimate, the FTC compared the number of uninsured motorist claims relative to other claims filed during 1996 to 2003 (*i.e.*, when credit-based insurance scores were becoming more widely used) for states in which these scores were used and in states in which they were not.

We assessed how often consumers filed uninsured motorist claims relative to how often they filed bodily injury claims and property damage liability claims. Figure 6 shows that the number of uninsured motorist claims filed compared to the number of bodily injury claims filed increased in states where credit-based insurance scores were allowed, but decreased slightly in states where they were not.[76] Figure 6 also shows the number of uninsured motorist claims filed compared to the number of property damage claims filed was basically unchanged in states where scores were allowed and decreased somewhat in states where they were not.

These results are consistent with the hypothesis that scores, because they raise the premiums of some consumers, cause a larger share of consumers to drive without insurance[77] and/or more risky consumers to drive without insurance.[78] These results,

---

because – if uninsured – the driver would have to pay out less to cover others' losses from an accident than would the insurance company if the drivers bought insurance.

[76] The states identified as not allowing the use of scores during the relevant period of time are California, New Jersey, Massachusetts, and Hawaii. Because of limitations in the data, Texas and South Carolina are not included in either group.

[77] If reduced cross-subsidization leads to more consumers driving without insurance, this could actually lead to lower overall losses from accidents. Research shows that the effect on accidents of requiring drivers to buy liability insurance in order to operate a car legally is unclear. Some drivers may choose not to purchase insurance and then either not drive or drive less often or more carefully, to avoid detection, leading to fewer accidents. Other drivers may purchase insurance they otherwise would have foregone, and then drive more often or more riskily, because they no longer bear the liability risk of causing an accident, (continued)

however, should be treated with caution. First, the relative change between the groups of states took place during the period 1997 – 2000. While scores were becoming widely used during this period, credit-based insurance scoring had probably not yet affected most consumers' premiums, given that insurance companies generally do not use scores when renewing customers. Perhaps more importantly, the FTC's analysis could be affected by any state-specific changes in insurance markets. Because the number of states not allowing the use of credit-based insurance scores for automobile insurance was small (California, Hawaii, Massachusetts, and New Jersey), any such changes could render them unreliable as a comparison group. In addition, because the analysis relies on uninsured motorist claims to indirectly measure the level of driving without insurance, differences over time in which consumers carried uninsured motorist coverage in states which allow the use of scores and those that do not could affect the results.

3. Adverse Selection

Credit-based insurance scores also may make insurance markets more efficient if they decrease the extent to which consumers make insurance purchasing decisions using better risk information than that available to insurance companies. In a competitive market, insurers will offer prices to groups of consumers reflecting the average expected risk of loss for each group. But if a consumer has better information than the insurance

---

leading to more accidents. One study has reported that the latter effect predominates over the former effect. Alma Cohen and Rajeev Dehejia, *The Effect of Automobile Insurance and Accident Liability Laws on Traffic Fatalities*, 47 J. OF LAW AND ECON. 357 (Oct. 2004).

[78] If these results do reflect effects of credit-based insurance scores, they could have the indirect effect of mitigating some of the savings that higher-score drivers get from the use of scores. If more higher-risk drivers are uninsured, this could increase the expected cost of uninsured motorist claims that insured drivers submit under their policies. In turn, this could increase the premiums that insurance companies must charge lower-risk consumers to cover these increased uninsured driver claims. Accordingly, even these increases in the cost of uninsured motorist coverage may offset somewhat the decrease in premiums that higher-score consumers receive from the use of scores.

company about his or her own true expected risk of loss, he or she will know whether the group price is higher or lower than his or her true expected claims cost. The consumer may use this superior knowledge to determine whether and how much insurance to purchase, a phenomenon known as "adverse selection."[79]

Adverse selection may be occurring if higher-risk consumers are more likely to have insurance or more complete coverage than lower-risk consumers. A higher-risk consumer who realizes that he or she is being charged a price that is lower than his or her actual risk of loss cost will have an incentive to purchase more insurance coverage.[80] A lower-risk consumer who realizes that he or she is being charged a price higher than his or her actual risk of loss will have an incentive to purchase less insurance coverage.[81] If higher-risk consumers purchase more insurance coverage and lower-risk consumers purchase less insurance coverage, the average risk of the group of consumers who do buy insurance will be higher. Premiums then would have to increase for insurance companies to cover the total claims costs, providing a further disincentive for lower-risk consumers to purchase insurance. If consumers know more about the risk of loss they pose than insurance companies, it therefore can affect insurance purchasing decisions in ways that cause economic inefficiency.[82]

---

[79] The discussion here is of market-wide adverse selection, where consumers know more about their risk than any firm does. A firm competing in an insurance market can face another form of adverse selection if one of its competitors is able to do a better job of predicting risk and entices away low-risk customers while leaving behind high-risk customers. *See* State Farm Comment, *supra* note 66, at 8.

[80] It is conventional wisdom in the insurance industry, however, that the riskiest drivers are those who choose to buy the least amount of coverage possible, and would buy no insurance if it were not legally required. This conventional wisdom probably reflects, at least in part, that the "riskiest drivers" in question are riskiest based on characteristics that are used to underwrite and rate policies, like driving history, and they are charged the highest rates. If these drivers really are very risky, simple theory would predict that they would still be willing to pay very high rates. Explanations for why these drivers would be unwilling to buy insurance at rates that reflect their true risks could include: that the drivers have very limited assets and are therefore "judgment proof," and therefore face less actual risk than the firm would face; that these drivers are less risk-averse than other drivers, or even risk-loving, and therefore unwilling to buy insurance at market rates; that the drivers believe themselves to be less risky than firms judge them to be; or that the drivers are cash-constrained, and do not buy insurance even though they would rather have the insurance than face the risk of a large loss.

[81] A consumer may do this by purchasing a policy with large deductibles or low liability limits, by not purchasing certain types of coverage, or by not purchasing insurance at all.

If scores allow insurance companies to predict risk more accurately, it could decrease the difference between what consumers and insurance companies know about the risk that individual consumers pose. Insurance companies therefore would be able to charge consumers premiums that more accurately reflect the true risk. This would reduce the incentive of higher-risk consumers to purchase more insurance and lower-risk consumers to purchase less. Accordingly, scores may reduce the extent of adverse selection and make insurance markets more efficient.[83]

The FTC considered whether adverse selection exists in automobile insurance markets in the United States.[84] It seems unlikely that consumers have better information about the risk they pose than do insurance companies. Although consumers might have some sense of how much risk they pose based on their own experience, it seems unlikely that this sense is more accurate than the assessment insurance companies can make.

---

[82] Insurers who realize that adverse coverage selection is occurring may attempt to separate the higher- from the lower-risk consumers by offering different price-coverage combinations. One theoretical analysis suggests that under some conditions, this approach can reduce the inefficiency caused by adverse selection. Michael Rothschild and Joseph Stiglitz, *Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*, 90 THE Q. J. OF ECON. 629 (Nov. 1976). In fact, firms do offer different deductible choices, which could be a mechanism for separating high-risk and low-risk customers. It may also be a pricing response by firms to differing levels of risk-aversion among customers.

[83] The American Insurance Association has stated that an insurance company found that, after introducing the use of credit-based insurance scores, "there is some evidence that higher limits of liability coverage *and* lower physical damage deductibles are being purchased…" AIA Comment, *supra* note 66, at 14 (emphasis in original). This would be consistent with a reduction in adverse selection resulting from the use of scores.

[84] Empirical studies have found only limited evidence for adverse selection in automobile markets in other countries, and even then only in very special circumstances. *See* Alma Cohen, *Asymmetrical Information and Learning: Evidence from the Automobile Insurance Market*, 87 REV. OF ECON. AND STATS. 197 (May 2005) (finding evidence that lower risk drivers in Israel purchased less insurance coverage than higher risk drivers, but only for experienced drivers for a limited period of time after switching policies and in a country in which insurance companies do not share data on prior claims); Pierre-Andre Chiappori, et al., *Asymmetrical Information in Insurance: General Testable Implications*, (Feb. 24, 2004) (finding no evidence that lower risk drivers in France bought less insurance than higher risk drivers), *available at* http://www.iue.it/FinConsEU/papers2004/salanie.pdf.

Specifically, companies make their assessment also using information about the consumer's past experience, such as extensive prior claim information included in a database that insurance companies share and public record information, such as convictions for driving while intoxicated or speeding.

Moreover, even assuming that consumers have better knowledge than insurance companies of the risk they pose, there are significant limitations on the extent to which consumers can use this advantage to alter their insurance purchasing decisions. Most states mandate minimum liability coverage for cars, and lenders typically require even greater coverage on cars they finance. Even though consumers retain some ability to make choices concerning insurance coverage, such as deductibles and limits, these choices are limited considerably.[85]

The FTC analyzed its automobile insurance database to test whether there was any indication that adverse selection may be occurring. We found that lower-risk drivers tend to have policies with higher deductibles than do higher-risk drivers, that is, lower-risk drivers have less insurance coverage than higher-risk drivers. This is consistent with (but does not prove) adverse selection is occurring in automobile insurance markets.[86]

If credit-based insurance score information is considered in the analysis, *i.e.*, the risk information available to insurance companies relative to consumers is enhanced, then adverse selection would be expected to decrease. However, when the FTC considered scores in its analysis, lower-risk drivers were still found to have insurance policies with higher deductibles than higher-risk drivers. This suggests that adverse selection may not be occurring, or, if it is occurring, then scores may not reduce it.

---

[85] It is clear that adverse selection experienced by a single firm can be a powerful force. When different firms have significantly different risk prediction technology, consumers will see the different prices charged and will tend to choose the firms with lower premiums. This can lead to a negative-feedback loop that can even cause a firm to collapse.

[86] One alternative explanation is moral hazard. If people with more complete coverage take less care, because they bear less of the cost of any accident or other damage or loss, this would result in the same relationship.

B. Other Possible Effects of Credit-Based Insurance Scores

Innovations in risk prediction techniques like credit-based insurance scores may affect the availability of insurance and some of the costs associated with selling insurance. First, some consumers may have a broader range of options to choose from when purchasing insurance. Because credit-based insurance scores predict risk more accurately for consumers, insurance companies now may be willing to offer coverage to some higher-risk consumers. In addition, credit-based insurance scores may make the process of underwriting and rating quicker and cheaper, and competition between insurance companies may cause cost savings from these process improvements to be passed on to consumers in the form of lower premiums.

Insurance companies and industry representatives stated that the use of credit-based insurance scores gives firms greater confidence in their ability to predict the risk that consumers pose. That is, if firms have more confidence in their risk estimates, they may be able to offer insurance to customers for whom they would otherwise not be able to determine an appropriate premium. The American Insurance Association, for example, has stated "(m)ore precise pricing enables insurers to accept greater risk by ensuring that both good risks and more marginal risks are properly priced to reflect the exposure they represent."[87] Several firms, including The Hartford and MetLife Home and Auto, have stated that the use of credit-based risk scores enabled them to offer policies to higher-risk consumers than they had previously.[88] This could lead to higher-risk consumers having more choices as they shop for insurance. No data, however, were submitted or obtained to assess the extent to which credit-based insurance scores actually have expanded insurance choices for higher-risk consumers.

---

[87] AIA Comment, *supra* note 66, at 4.

In addition, several insurance companies and score developers emphasized that the use of scores can save costs. Specifically, they asserted that the use of scores facilitates automation, speeds up policy underwriting and rating, and otherwise reduces the costs of underwriting and rating.[89] No data was submitted or obtained to allow the FTC to develop reliable estimates of cost-savings associated with credit-based insurance scores. Assuming that there are such savings, the FTC would anticipate that competition in the market for automobile insurance would result in these savings being passed on to consumers in the form of lower prices.

Further, banning the use of factors that are known to be correlated with risk could have negative effects on insurance markets. If firms cannot adjust prices based on the risk associated with a characteristic, they will have an incentive to refuse to offer policies to people with the characteristic.[90] If the law prohibits firms from refusing to sell policies to people with that characteristic, they will still have an incentive to try to avoid insuring them. This could cause firms to expend resources on finding ways to avoid higher-risk consumers, reducing the availability of insurance to higher-risk consumers and making otherwise profitable distribution channels untenable.

---

[88] Meeting between FTC staff and The Hartford (July 14, 2004); Meeting between FTC staff and MetLife Home and Auto (July 12, 2004); Meeting between FTC staff and USAA, (July 14, 2004). *See also* AIA Comment, *supra* note 66, at 7-8; NAMIC Comment, *supra* note 20, at 6-7.

[89] AIA Comment, *supra* note 66, at 12; Fair Isaac Comment, *supra* note 18, at 15; State Farm Comment, *supra* note 66, at 4.

[90] *Id*. at 10.

A simple example illustrates the possible impact of banning the use of a characteristic in making the decision whether to offer insurance to consumers. Assume that geographic location is correlated with risk on insurance policies[91] and that firms are allowed to refuse to sell insurance based on geography but are not allowed to charge different prices based on geography. This would give insurance companies an incentive to refuse to sell policies to people living in riskier areas.[92] If firms could not outright refuse to sell policies based on geography, and could not charge different prices based on geography, they would have an incentive to use other means to avoid insuring those who live in more risky geographic areas, for example, not establishing offices, working with independent agents, or advertising in these locations.

It is not clear, however, whether banning the use of credit-based insurance scores would lead to distortions of the insurance market like those associated with banning the use of geography. An insurance company does not see a consumer's score until he or she applies for insurance coverage. It therefore would be difficult for insurance companies to directly avoid selling insurance to consumers with low scores. There may be, however, different marketing approaches, such as alternative types of advertising, which bring in consumers with different average scores. If firms cannot use scores to underwrite or rate, they would have an incentive to market only through channels that bring in consumers with higher scores. This could reduce the availability of information about insurance options, particularly to consumers with lower insurance scores. No data was submitted or obtained, however, to permit the FTC to determine whether restrictions on the use of scores actually would have this type of effect.

---

[91] *Id*.

[92] Firms might specialize geographically, with firms with higher premiums offering policies everywhere but mainly getting customers from high risk areas, while lower-cost firms refuse to write in high-risk areas.

Banning credit-based insurance scores may also give firms greater incentives to invest in developing other risk-prediction tools. If the use of scores is banned, firms may have an incentive to spend more on developing new risk variables to capture some of the same risk prediction benefits of scores.[93] This could be seen as an unnecessary societal cost, given that scoring technology has already been developed and scores are a fairly low-cost risk prediction technique.

C. Effects on Residual Markets for Automobile Insurance

The introduction and growth in the use of credit-based insurance scores has taken place during a time when one particular measure of the functioning of the market, the share of consumers buying insurance through state-run "residual markets," indicated the market was working well. All states run some type of program to allow consumers to purchase automobile insurance when they are unable to find a private firm willing to sell them policies voluntarily.[94] To avoid attracting consumers who could otherwise obtain private insurance coverage, these state-run programs typically charge higher prices than private insurance companies.

Figure 7 shows the share of automobile policies that were purchased through state-run programs during the years 1996 – 2003, broken down by states that allow the use of credit-based insurance scores, and those that do not.[95] It shows that a larger share of consumers participated in these programs in the states that did not allow the use of scores. However, this was true both before and after the introduction of scores, and therefore this difference in levels of participation presumably reflects other differences between states. Figure 7 shows that the state-run program share fell during the second half of the 1990s, as score were being introduced, and then leveled off after 2000. The pattern is nearly identical in states that allowed the use of scores and states that did not. Therefore, Figure 7 is probably best interpreted as meaning that scores at least did not interfere with the smooth functioning of automobile insurance markets.

---

[93] *See* Cheng-Sheng Peter Wu, Deloitte and Touche, *What to do When You Cannot Use Credit* (Sept. 2000) (presentation at the CAS Special Interest Seminar), *available at* http://www.casact.org/education/specsem/f2005/handouts/credit.ppt.

[94] *See* https://www.aipso.com/about.asp.

[95] The states identified as not allowing the use of scores during the relevant period of time are California, New Jersey, Massachusetts, and Hawaii. Because of limitations in the data, Texas and South Carolina are not included in either group.

VI. EFFECTS OF SCORES ON PROTECTED CLASSES OF CONSUMERS

FACTA requires that the FTC analyze the extent to which the use of credit-based insurance scores affects the availability and affordability of insurance for members of certain categories of consumers. The statute mandates that the Commission consider the impact of these scores on categories of consumers based on race, ethnicity, national origin, geography, income, religion, age, sex, and marital status. In particular, the agency was instructed to assess whether scores act as a proxy for membership in these groups.

In fulfilling the statutory mandate, the FTC focused its analysis on the effect of credit-based insurance scores on consumers in different racial, ethnic, national origin, and income groups. The Commission did not focus its assessment on consumers in different religious groups because we are not aware of any reliable data relating scores to religious affiliation. In addition, the FTC also did not focus its analysis on consumers in different geographic, age, sex, and marital status groups. In most locations in the United States, insurance companies can and do use geography, age, sex, and marital status directly in determining automobile insurance premiums.[96] While credit-based insurance scores may vary based on these factors, the direct effect of using these factors to price insurance far outweighs any indirect effects these factors may have through their impact on scores. The FTC therefore did not try to measure any such indirect effects.

A. Credit-Based Insurance Scores and Racial, Ethnic, and Income Groups

1. Difference in Scores across Groups

The FTC analyzed whether there was a relationship between credit-based insurance scores and race, ethnicity, national origin, and income. In undertaking this analysis, the Commission first reviewed and considered prior research. In 1999, the Virginia State Corporation Commission's Bureau of Insurance issued a report assessing the relationship at the ZIP code level between scores and race as well as between scores and income.[97] The report stated that "nothing in (our) analysis leads the Bureau to the conclusion that income or race alone is a reliable predictor of credit scores." Nevertheless, the absence of more detailed information about the results of this study leaves unclear the relationship between scores and race and income.

The State of Missouri Department of Insurance released a study in 2004 that relied on similar data.[98] The Missouri study used ZIP-code level data on scores and race, income, and other demographic variables. The scores analyzed were credit-based insurance scores that twelve large insurance companies used for automobile or homeowners policies. The Missouri study found that scores were correlated with the racial, ethnic, and income characteristics of ZIP codes. Specifically, as the proportion of racial and ethnic minorities or lower-income consumers in a ZIP code increased, scores decreased.[99] These correlations remained after controlling for education, marital status, and housing values.

---

[96] While the use of income to underwrite policies or set rates may not be expressly prohibited in some locations, it appears to be generally regarded as an illegitimate variable for those purposes.

[97] Report of the State Corp. Comm'n's Bureau of Ins. to the Sen. Commerce and Labor Comm. of the Gen. Assemb. of Va., *Use of Credit Reports in Underwriting* (1999) [hereinafter Virginia Study].

[98] Brent Kabler, Ph.D., *Insurance-Based Credit Scores: Impact on Minority and Low Income Populations in Missouri* (Jan. 2004) [hereinafter Missouri Study], *available at* http://www.insurance.mo.gov/reports/credscore.pdf.

Unlike prior researchers, the Texas Department of Insurance (TDI) in its 2004 study moved beyond aggregate data and obtained data about individuals to analyze the relationship between scores and race, ethnicity, and income. The TDI used automobile and homeowners policy data from six large insurance companies. The TDI obtained race data for each consumer from the Texas Department of Public Safety and ethnicity data for each consumer from a Hispanic surname match. The TDI used median income for the ZIP code in which consumers lived, because individual income information was not available.

The TDI's analysis of this data showed that African Americans and Hispanics tended to have lower scores than Asians and whites.[100] It revealed mixed results for income. For some insurance companies, consumers in higher-income areas had higher scores, while this was not the case for other insurance companies. It is not clear whether these different results for income reflect differences in the credit scoring models that the insurance companies used, or in the mix of customers at each firm.

---

[99] An attempt was also made in the Missouri study to use the ZIP code level data to draw inferences about individual-level differences in credit scores by race and income. The results of this analysis were more speculative, but did demonstrate that it would be very unlikely that the differences found at the ZIP code level could have been found if there were no differences at the individual level.

[100] The TDI characterized the scores in this way: "In general, Blacks have an average credit score that is roughly 10% to 35% worse than the credit scores for Whites. Hispanics have an average credit score that is roughly 5% to 25% worse than those for Whites. Asians have average credit scores that are about the same or slightly worse than those for Whites." 2004 Texas Report, *supra* note 41, at 13.

After reviewing the prior research, the FTC analyzed the information in its own automobile insurance database to assess the relationship between scores and race, ethnicity, national origin, and income. Figure 8 shows how non-Hispanic whites, African Americans, Hispanics, and Asians are distributed across the range of scores. The horizontal axis shows score deciles, and the vertical axis shows the share of each group that falls in each decile. The deciles are defined using the overall distribution of scores. If a group had the same distribution of scores as the overall sample, then 10% of that group's population therefore would fall in each of the ten deciles.

Figure 8 shows that non-Hispanic whites and Asians are fairly evenly distributed across the score range, resulting in a roughly flat line near 10%. In contrast, African Americans and Hispanics are strongly over-represented in the lowest deciles and under-represented in the highest deciles. For example, 26% of African Americans are in the group with the lowest 10% of credit-based insurance scores, while only 3% are in the highest 10% of scores. Similarly, 19% of Hispanics are in the group with the lowest 10% of scores, and 5% are in the highest 10% of scores.

Another way of measuring these differences is to look at where the median person[101] for each racial or ethnic group falls in the overall distribution of scores. If scores were distributed identically across racial and ethnic groups, the median score for each group would equal the overall median – the 50th percentile. The FTC's data show that the median scores for non-Hispanic whites and Asians are quite similar to that of the overall sample, with the median score for non-Hispanic whites and Asians falling in the 54th and 52nd percentile, respectively. In contrast, the median scores for the African

---

[101] One-half of the group will have a score lower than the median person and one-half will have a score higher than the median person.

Americans and Hispanics are much lower, with the median scores of African Americans and Hispanics falling in the 23rd and 32nd percentile, respectively. So, more than one-half of all African Americans have credit scores in the lowest quarter of the overall score distribution, and one-half of all Hispanics have credit scores in the lowest third of the overall score distribution.

Figure 9 presents an alternative way of viewing these differences. It shows the racial and ethnic makeup of each decile in the score distribution, which varies considerably across the range of scores. Because non-Hispanic whites make up such a large share of the populations, they are a majority in every score decile. But, as Figure 8 shows, African Americans and Hispanics are heavily over-represented in the lower score deciles.

In addition to race and ethnicity, the FTC examined the relationship between scores and national origin. To assess this relationship, the Commission compared scores for foreign-born consumers with those of consumers born in the United States. The scores for consumers born outside the United States were slightly lower than those of consumers born in the United States, with the median score of the foreign-born consumers falling in the 44th percentile of all scores.

The FTC also compared the scores for recent immigrants and other consumers.[102] Recent immigrants have scores that are slightly lower than other immigrants and lower than consumers overall, with the median score for recent immigrants falling in the 39th percentile of all scores. We found that recent immigrants whose information is included in the FTC database were much more likely to be Hispanic or Asian than consumers born in the United States. This makes it complicated to evaluate and describe the relationship between scores and race or ethnicity apart from the effect of national origin. Because race and ethnicity are associated with much larger differences in scores than national origin, the Commission focused its further analysis on race and ethnicity.

---

[102] The FTC database does not contain information on the actual date of anyone's arrival in the United States. For this reason, recent immigrants were defined as people who first applied for a Social Security card during the previous ten years, and who were 30 years old or older at the time of the sample. These restrictions were an attempt to limit "recent immigrants" to people who arrived in the United States as adults.

Finally, the FTC study evaluated the relationship between scores and income. The Commission did not have access to information about the income of the particular consumers in its database. The FTC instead used the median income of the United States Census tract in which consumers live to divide them into low-to-moderate income, middle income, and high-income neighborhood groups.[103] Figure 10 shows the share of people in each income category in each decile of the distribution of scores. Low-to-moderate income consumers are somewhat over-represented in the lower score deciles, with 15% of these individuals in the lowest 10% of scores, and only 8% in the highest 10% of scores. Middle-income consumers are essentially evenly distributed across the distribution of scores. High-income consumers are under-represented in the lowest 10% of the score distribution, but otherwise fairly evenly distributed. Figure 11 shows the income breakdown of each score decile. Again, it shows that there is some relationship between neighborhood income and score.

The results for the FTC's database show that as income increases, scores tend to increase. These results, however, are much weaker than the results for race and ethnicity.

---

[103] This approach follows methods used to analyze income in FRB studies of mortgage markets. The groups were: (1) Low-to-moderate income: Tract median < 80% of MSA median income; (2) Middle income: Tract median >= 80% of MSA median income and <%120 of MSA median income; and (3) High income: Tract median >= %120 of MSA median income. As discussed in Appendix F, we have also done much of the analysis using absolute median income, instead of income relative to the MSA, and the results are not qualitatively different.

This may be because the relationship between score and income actually is weaker, or it may simply be the result of only having data on income at the neighborhood level.

2. Possible Reasons for Differences in Scores across Groups

As discussed above, the FTC's analysis shows a relationship between credit-based insurance scores and race, ethnicity, and, to a lesser extent, income. The Commission examined other information in its sample to determine what factors could account for differences in scores among racial and ethnic groups. The FTC's database contains some information on factors that could explain some of the differences in scores among racial and ethnic groups. Specifically, it includes information on the median income of the neighborhood in which each consumer lives, and consumers who live in lower-income neighborhoods tend to have lower scores. It also contains information from which the age of the consumers whose score is in the database can be inferred,[104] and older consumers tend to have higher scores. Finally, the FTC's database contains information about the gender of the consumers whose score is included (the "first named insured" on the policy), and men in the FTC database tend to have higher scores than women, although the difference in average score between men and women in the FTC database cannot be generalized to the overall population.[105]

---

[104] For single-driver households, we know the age of the person for whom we have a credit score. For multi-driver households, we need to make an assumption about whose age we have. We do this in several ways. From Social Security Administration ("SSA") data, we know the gender of the person whose credit score we have. If there is only one driver in a household with that same gender, we assume that person is the person for whom we have a credit score. If there are multiple people whose gender matches the SSA data, we take the oldest, on the assumption that that person is most likely to be the first named insured.

[105] We have a score for only one person covered by each policy. From examining our data, it is apparent that in households with male and female adults (*e.g.*, married couples), it is most often the male driver who is the first named insured, and therefore the person for whom we have a score. About 75% of multi-driver policies have a male first named insured, while the split for single-driver policies is 50/50. So, it appears that the men for whom we have scores are much more likely to be married than the women for whom we

(continued)

Table 3 presents median neighborhood income, age, and gender for racial and ethnic groups for consumers whose information is in the FTC database. It shows that African Americans and Hispanics live in neighborhoods with lower median incomes than non-Hispanic whites and Asians. It reveals that Hispanics and Asians are younger than non-Hispanic whites and African Americans. It further shows that the African-American customers in this sample are much more likely to be female than are customers in other racial and ethnic groups.[106] All of these differences are consistent with African Americans and Hispanics having lower credit scores.

Figure 12 shows the distribution of scores by race and ethnicity after controlling for neighborhood income, age, and gender of the person scored. It shows that large differences remain in the distributions of scores across racial and ethnic groups, and that these differences are only slightly smaller than they were prior to controlling for these factors.[107] In particular, prior to controlling for these factors, the median score for African Americans and Hispanics was in the 23rd and 32nd percentiles, respectively. After using these controls, the median score for African Americans and Hispanics rose to the 27th and 37th percentiles, respectively. In short, consideration of neighborhood income, age, and gender explains only a small part of the difference in credit-based insurance scores between racial and ethnic groups. It is not clear what explains the rest of the difference.

---

have scores. The differences in score by gender in the FTC database, therefore, cannot be interpreted as the difference in scores that would be observed between all men and all women, because they also reflect differences in credit score by marital status and household size.

[106] Recall that age and gender, like score, are for the customer who was the “first named insured” on a policy.

[107] It is our understanding that the Federal Reserve Board is undertaking a similar analysis using a richer set of data about each individual.

3. Impact of Differences in Scores on Premiums Paid

a. Effect on Those for Whom Scores Were Available

The FTC assessed the implications of the differences in credit-based insurance scores for the premiums that members of different racial and ethnic groups would be predicted to pay. As discussed above, the FTC database can be used to predict differences in claims risk with and without the use of scores. These differences, in turn, can be used to estimate the effects of scores on expected insurance premiums for racial, ethnic, and income groups.

Figure 13 shows the results of the FTC's analysis. These are graphs that show the share of each group with different size changes in their predicted risk between models where scores were not used and models where scores were used. Comparing across groups clearly shows that a much larger share of African American and Hispanics had increases in their predicted risk than did non-Hispanic whites and Asians. When scores are used, the predicted risk decreased for 62% of non-Hispanic whites and 66% of Asians. On the other hand, the predicted risk increased for 64% of African Americans and 53% of Hispanics. These results flow from the fact that, as discussed above, the scores for African Americans and Hispanics are lower on average than the scores of non-Hispanic whites and Asians.

Table 4 shows the magnitude of changes in predicted risk for racial and ethnic groups as a result of the use of scores. The average predicted risk increased by 10% for

African Americans and 4.2% for Hispanics, and dropped by 1.6% for non-Hispanic whites and 4.9% for Asians.[108]

b. Effect on Those for Whom Scores Were Not Available

The FTC also sought to determine whether the likelihood that a credit-based insurance score could not be generated for a consumer varied across racial and ethnic groups, and what impact any such differences would be expected to have on the premiums paid by consumers. A score may not be available for a consumer for one of two reasons: either it cannot be located for a consumer (a "no-hit"), or a consumer may have a credit history file, but it may not contain information sufficient to calculate a credit-based insurance score (a "thin file").

The FTC database does not contain Social Security Administration race and ethnicity data for most customers who were "no hits" or "thin files."[109] The FTC therefore used United States Census data to determine whether there are differences in the proportions of racial and ethnic groups that do not have a credit score. Based on block-level data, the Commission estimates that credit reports could not be located for 9.7% of African Americans, 9.2% of Hispanics, 7.8% of non-Hispanic whites, and 6.4% of Asians. Similarly, 2.4% of Hispanics, 2.1% of African Americans, 1.8% of non-

---

[108] The relatively large decrease in predicted risk for Asians relative to non-Hispanic whites was surprising, given how similar the score distributions are for these two groups. In addition, the increase in predicted risk for Hispanics was only half that of African Americans, even though Hispanics have average scores closer to African American than to the overall population. Further examination of the results of the models showed that the inclusion of scores affected the impact of other variables on predicted risk. This, in turn, affected the predicted risk of Asians and Hispanics. In particular, the impact that short tenure with a firm and low liability limits had on predicted risk shrank considerably when scores were included in the models. Asians and Hispanics have low average tenure and low average liability limits, so when the impact of those characteristics on predicted risk decreased, so did the average predicted risk of Asians and Hispanics.

[109] The process of obtaining SSA race and ethnicity data relied on obtaining Social Security Numbers or dates of birth from credit reports; thus we did not receive SSA information for people whose credit reports could not be located, or who had very little information in their reports. Similarly, we do not have SSA national origin information for these people, and therefore cannot analyze the impact on immigrants of a lack of a credit-based insurance score.

Hispanic whites, and 1.8% of Asians had credit reports, but with too little information available to calculate a score.[110] Note that because these results are based on geographic data, they may not exactly reflect actual differences between racial and ethnic groups.

The FTC's assessment indicates that consumers for whom scores were not available appeared slightly riskier when scores were considered than when they were not. The Commission compared the results from risk models without scores with results from risk models with scores that also included categories for "no hit" and "thin file" in making this determination. No-hit consumers were 1.06 times riskier in a model that included controls for scores compared to a model that did not. Thin-file consumers were 1.02 times riskier in a model that included controls for scores compared to a model that did not.

Given the relatively small differences across groups in the share of people who were "no hits" or "thin files," and the relatively small impact of not having a score on predicted risk (as opposed to the large impacts of using scores on the predicted risk of people in the lowest score deciles, for example), this is unlikely to be an important source of differences in premiums across racial and ethnic groups. Again, this analysis is limited by the lack of individual-level data on race and ethnicity for people for whom we do not have credit scores.

---

[110] The block data were used by assuming each person had a likelihood of being a member of each racial or ethnic group that was proportional to the share of the population of each group in that person's block. This is implemented similarly to how imputed race/ethnicity information for SSA data are used. See Appendix C for a discussion of that process.

B. Scores as a Proxy for Race and Ethnicity

Section 215 of FACTA mandates that the FTC create a statistical model of insurance claims that includes credit-based insurance scores, standard non-credit risk variables, and controls for protected classes under the Equal Credit Opportunity Act.[111] We understand this to require the agency to analyze whether credit-based insurance scores act as a "proxy" for membership in these classes. As discussed above, we focused our analysis on effects on different groups defined by race, ethnicity, and income.

Understanding how a proxy functions is critical to the FTC's analysis. Insurance companies build statistical models that relate a variety of characteristics of customers (*e.g.*, age or driving history) to risk. Firms then use these models to predict the average claims that customers with those characteristics will generate, and these predictions of risk play a central role in determining the premiums that firms charge.

The risk models that companies build do not include information about race, ethnicity, or income. If there are large differences in average risk based on race, ethnicity, or income, then models may attribute some of those differences in risk to other variables included in the model that differ across these groups. The included variable thus may act in whole or in part as a statistical "proxy" for the excluded variables of race, ethnicity, or income.[112]

The FTC sought to determine whether credit-based insurance scores act as a proxy for race, ethnicity, and income in insurance decisions. To determine whether there is such an effect, and, if so, its magnitude, the Commission conducted three related analyses. First, the Commission analyzed whether scores predict risk within racial, ethnic, and income groups. If scores do not predict risk within any group defined by race, ethnicity, and income, then the sole reason that scores predict risk in the general population would be because they act as a proxy for membership in different groups.

---

[111] FACTA § 215(a)(2) (2006); 15 U.S.C. §1681 note (2006).

[112] The econometric term for this effect is "omitted variable bias." The omission of a predictive variable (such as race, ethnicity, or income) causes the estimated effect of a variable that is correlated with the omitted variables, such as score, to be "biased" away from the true effect. In this scenario, the direction of the bias would be to overstate the relationship between score and claims.

Second, the Commission analyzed whether average risk differed substantially by race, ethnicity, and income. If there were no substantial differences in the average risk across racial and ethnic groups, then there would be no underlying difference for which scores could act as a proxy. If there are substantial differences in risk across groups, scores may in part act as a proxy, even if scores also predict risk within groups (and are therefore not solely acting as a proxy for membership in different groups).

Third, the FTC created models that included controls for race, ethnicity, and income, along with credit-based insurance scores and the full range of other predictive variables. The Commission quantified the proxy effect of scores by measuring the impact of including these additional controls on the estimated relationship between scores and claims. To provide a basis for comparison, the FTC also conducted this analysis for several other variables that are predictive of risk.

1. Do Scores Act Solely as a Proxy for Race, Ethnicity, or Income?

Whether credit-based insurance scores predict risk within racial, ethnic, and income groups provides critical insight into whether scores are a proxy for membership in these groups. If scores did not predict claims within racial, ethnic, and income groups, the relationship between scores and claims must come from scores acting as a proxy for race, ethnicity, and income. On the other hand, if scores do predict risk within groups, then they do not serve solely as a proxy if used to assess risk for all consumers.

Therefore, the FTC analyzed whether scores predict risk within race, ethnicity, and income groups.

The results of the FTC's analysis are presented in Figure 14 for each racial and ethnic group for each type of automobile insurance coverage. If credit-based insurance scores predict the amount that insurance companies paid out in claims within each group, there should be a downward slope on each graph.[113] In other words, as scores increase for members of each group, the amount paid out on claims should be decreasing.

Although the relatively small sample size for the minority groups in the FTC database (which is a particular problem for bodily injury coverage, which has relatively few claims) leads to results that sometimes vary substantially from decile to decile, the overall pattern observed is that the amount paid out decreases as credit-based insurance scores increase for each group for each type of coverage.[114] With the exception of collision coverage, very few of the decile and coverage combinations have estimated risk for a given racial or ethnic group that is statistically significantly different from that of the overall sample.[115] Because they show that scores predict risk within groups, these results show that credit-based insurance scores do not predict risk solely by acting as a proxy for membership in racial and ethnic groups.[116]

---

[113] These were estimated by including interaction terms between the race/ethnicity variable and the scores variables. The coefficients on non-race/ethnicity non-score variables are therefore forced to be the same across groups. Entirely separate models cannot be estimated for many race/ethnicity/coverage combinations, because the small sample size of the minority groups often leads to the non-convergence of the estimation procedure.

[114] One cell that jumps out as being out of line with that pattern is the ninth decile for African Americans for comprehensive coverage. Further investigation showed that this result was affected by an outlier; a single individual with a very large claim, very low earned car years, and a very large nationally-representative weight had a large impact on the estimated risk for this decile. There are also few African Americans in the ninth decile. The difference between the estimated risk for African Americans in the ninth decile and the overall sample in the ninth decile was not statistically significant. When this outlier was dropped the risk estimate for this decile was similar to the surrounding deciles. The treatment of outliers is discussed in Appendix F.

[115] Statistical significance was determined using a bootstrap procedure with 500 replications. The bootstrap procedure is discussed in Appendix D.

The FTC conducted the same analysis based on neighborhood income. These results are shown in Figure 15. These graphs show a consistent negative relationship between amount paid and credit-based insurance score for all neighborhood income groups. In other words, as scores increased, claims decreased for all income groups.

In short, because scores do predict risk within racial, ethnic, and income groups, they do not act solely as a proxy for these characteristics.

2. Differences in Average Risk by Race, Ethnicity, and Income

Even though scores do not act solely as a proxy for race, ethnicity, and income, there may still be some proxy effect. For such a partial proxy effect to occur, there must be differences in average risk among racial, ethnic, or income groups, *i.e.*, scores can only have a proxy effect if there is an underlying relationship for which scores can serve as a proxy. To determine whether such differences exist, the FTC created models that evaluated the relative amount paid on claims by race, ethnicity, and neighborhood income for the four main types of automobile insurance coverage. These models included other risk variables, but not scores. The results of this analysis are shown in Table 5. For purposes of comparisons in these tables, the FTC assigned a relative value of 1 to the amount of claims that would be expected to be paid to non-Hispanic white consumers and to consumers living in high income neighborhoods.

Column (a) shows that Asians and Hispanics had a higher amount of claims paid under property damage liability coverage than did African Americans and non-Hispanic whites, although the difference was not statistically significant for Hispanics. It also shows that there was very little relationship between the amount of property damage liability claims and whether a consumer lives in a neighborhood with a low, middle, or high income. While Asians did have more claims under property damage liability coverage, as discussed above, our analysis showed that they had scores that were similar to the scores of the overall distribution. Therefore, scores cannot act as proxy for being Asian, so it is unlikely that scores could act as a proxy for race or ethnicity in a model of property damage liability claims.

---

[116] We also did the same analysis for “foreign born” and “recent immigrants.” The results were similar, and scores are correlated with risk for those groups.

Columns (b) through (d) of Table 5 present results concerning the amount paid out for bodily injury, collision, and comprehensive coverage, respectively. After controlling for other risk factors, insurance companies paid out 48% more to African Americans than non-Hispanic whites for bodily injury, 43% more for collision, and 63% more for comprehensive coverage. Similarly, they paid out 25% more to Hispanics than non-Hispanic whites for bodily injury, 33% more for collision, and 45% more for comprehensive coverage. Insurance companies paid out 30% more to Asians than non-Hispanic whites for collision coverage. These differences were all statistically significant. The differences for bodily injury liability and comprehensive coverages between Asians and non-Hispanic whites were relatively small and not statistically significant. The large differences in average risk on comprehensive coverage for Hispanics and African Americans should be treated with some caution, as the geographic risk variable in the FTC database is not a very effective control for geographic variation in risk on comprehensive coverage.[117]

Table 5 shows that the differences among neighborhood income groups were much smaller than those among racial and ethnic groups. The one substantial difference in risk was that customers in low-income neighborhoods pose a 16% higher risk for comprehensive coverage. Again, this may in part be due to the lack of an effective geographic risk measure for comprehensive coverage.[118]

These results show that there were substantial differences in the average risk of consumers in different racial and ethnic groups for all four major automobile insurance coverages.[119] For property damage liability coverage, Asians were the only group with

---

[117] The geographic risk measure in the FTC database is based on property damage liability claims, which result from accidents. The estimated effect of the geographic risk measure is much smaller in the comprehensive coverage risk models than in the models for the other coverages, suggesting that it is a poor control for geographic variation in comprehensive coverage risk. According to the Bureau of Justice Statistics, African Americans and Hispanics are much more likely to be victims of automobile theft (a risk covered by comprehensive coverage) than non-Hispanic whites. *See* Bureau of Justice Statistics file cv0516.csv, *available at* www.ojp.usdoj.gov/bjs/pub/sheets/cvus/2005/cv0516.csv; Bureau of Justice Statistics file cv0517.csv, *available at* www.ojp.usdoj.gov/bjs/pub/sheets/cvus/2005/cv0517.csv. In the absence of a good measure of the geographic variation in comprehensive coverage risk, race, ethnicity, and neighborhood income are likely picking up some of that variation in risk (*e.g.*, they may be acting as a proxy for other characteristics of neighborhoods that affect comprehensive coverage risk). Additional support for this hypothesis was found by estimating separate risk and severity models that included race, ethnicity, and income controls. In those models, race, ethnicity, and income affected only frequency in the property damage liability, bodily injury liability, and collision coverage models. In the comprehensive coverage model, race, ethnicity, and income were strongly related to claim severity. This is consistent with those variables being related to the likelihood of theft claims.

[118] *Id.*

[119] We found similar patterns when we used loss ratios as the measure of relative risk, instead of the direct results of the risk models. The loss ratio is the ratio of payments companies made on claims divided by premiums customers paid in. Using loss ratios, therefore, shows whether customers in different racial and ethnic groups generated greater or lesser total payouts on claims, on average, than predicted by the companies, as reflected in the premiums the customers were charged. Loss ratios were fairly similar across groups for property damage liability coverage, with Hispanics and Asians generating somewhat more claims relative to premiums than African Americans and non-Hispanic whites. For bodily injury liability coverage, collision coverage, and comprehensive coverage, African Americans and Hispanics generated higher claims relative to premiums than did non-Hispanic whites. The same was true for Asians for collision coverage, although Asians had a substantially smaller loss ratio for comprehensive coverage than did any other group. For example, the loss ratios of African Americans and Hispanics for collision coverage were 83.9% and 85.6%, respectively, for Asians 78.2%, and for non-Hispanic whites the loss ratio was 63.3%. Unlike in our risk models, the coverage with the largest differences across groups was bodily injury liability coverage, as opposed to comprehensive coverage. This again suggests that part of the reason we find such large differences in risk across groups for comprehensive coverage in our models is the lack of a geographic risk measure that is specific to risk on comprehensive coverage. For the four

(continued)

significantly higher risk. For the other three coverages, Hispanics and African Americans had substantially higher average payouts on claims than did non-Hispanic whites. Given that Hispanics and African Americans have much lower credit-based insurance scores, on average, than do non-Hispanic whites, there is the potential that scores could gain additional predictive power by acting as a proxy for race and ethnicity in models of claims under bodily injury, collision, and comprehensive coverages.

3. Controlling for Race, Ethnicity, and Income to Test for a Proxy Effect

a. Existence of a Proxy Effect

The FTC created models that evaluated the relative amount paid on claims by score decile with and without controls for membership in racial, ethnic, and income groups for the four main types of automobile insurance coverage. Table 6 shows the results.[120] For purposes of comparisons on this Table, the FTC assigned the relative value of 1 to: (1) the amount of claims that would be expected to be paid to consumers in the highest 10% of credit-based insurance scores; (2) non-Hispanic white consumers; and (3) consumers living in high-income Census tracts. For each coverage, the first column shows the predicted relative amount of claims for credit-based insurance score deciles for a model that does not include controls for race, ethnicity and income. The second column for each coverage shows the results from models that include scores and controls for the prohibited factors.

Comparing the two columns for property damage liability coverage (columns (a) and (b)) reveals that there was very little difference in the impact of credit-based

---

coverages combined, the loss ratios of the four groups were: for non-Hispanic whites, 62%; for African Americans, 80%; for Hispanics, 81%; and, for Asians, 67%.

[120] Again, the models used here are Tweedie GLMs. Modeling details are given in Appendix D.

insurance scores on predicted risk based on whether the model included controls for membership in a protected class. The only statistically significant difference was that the estimated relative risk for the lowest score decile was larger when protected class controls were included in the model.[121] This is opposite of the change that would occur if scores were acting as a proxy. This lack of a proxy effect is not surprising, given that the only statistically significant difference in risk by racial or ethnic group for this coverage was that Asians had higher average risk. As pointed out above, because Asians have similar scores, on average, as the population as a whole, scores cannot act as a proxy for being Asian. The lack of any proxy effect for property damage liability coverage is made very clear in Figure 16, which shows the estimated relationship between claims risk and credit-based insurance scores from Table 6.

Table 6 shows that the results were somewhat different for bodily injury liability, collision, and comprehensive coverage. These are the coverages for which African Americans and Hispanics had substantially higher average total payouts on claims than did non-Hispanic whites. The FTC's analysis revealed that including these controls did reduce somewhat the effect of scores on predicted risk for these three coverages. The results show, however, that scores do continue to predict claims strongly if controls for race, ethnicity, and income are included in the risk models, which means that scores do not predict risk primarily by acting as a proxy for these characteristics. In addition to Table 6, the results are presented in Figure 16, which shows the estimated relationship between scores and risk, with and without controls for race, ethnicity, and income. Controls for race, ethnicity, and income decreased the impact of scores on predicted risk for these coverages most for the lowest credit-based insurance score deciles (where African Americans and Hispanics are disproportionately located), and these decreases were statistically significant. In short, the FTC's analysis indicates that credit-based insurance scores appear to have some proxy effect for three of the four coverages studied, but that this is not the primary source of their relationship with claims risk. In the next section, we address the magnitude of the proxy effect.

---

[121] A 95% confidence interval for the difference between the score decile parameter estimates from the two models was computed using a bootstrap procedure with 500 replications. Details of the bootstrap procedure are provided in Appendix C.

b. Magnitude of a Proxy Effect

The FTC also sought to determine the magnitude of any proxy effect from the use of credit-based insurance scores. Controlling for race and ethnicity had the largest impact on the predicted effect of scores on risk for comprehensive coverage. *See* columns (g) and (h) of Table 6. Without these controls, consumers in the lowest 10% of scores were estimated to pose 1.95 times more risk than consumers in the highest 10%. With the controls, consumers in the lowest 10% of scores were estimated to pose 1.74 times more risk than consumers in the highest 10%. As discussed above, this result should be treated with caution, because it could be affected by the lack of a good measure of the geographic variation in comprehensive coverage risk.

Controlling for race and ethnicity had a smaller effect on the predicted impact of scores on risk for bodily injury liability and collision coverage. For bodily injury liability coverage, without these controls, consumers who are in the lowest 10% of credit-based insurance scores were estimated to pose 2.20 times more risk than consumers in the highest 10% of scores, while with controls they were estimated to pose only 2.10 times more risk. *See* columns (c) and (d) of Table 6. For collision coverage, without controls, consumers who are in the lowest 10% of credit-based insurance scores were estimated to pose 2.03 times more risk than consumers in the highest 10% of scores, while with controls they posed only 1.93 times more risk. *See* columns (e) and (f) of Table 6.

It may be difficult to interpret the magnitudes of the proxy effects by examining changes in the predicted effects of scores on claims risk. An alternative way to measure the magnitude of the proxy effect is to examine how it affects the impact of scores on the predicted risk of different race and ethnicity groups. The information presented in Table 7 compares the impact of scores on predicted risk for different groups, with and without race, ethnicity, and income controls. The first column in Table 7 shows that if scores were used, then on average the predicted risk of African Americans increased by 10% and Hispanics increased by 4.2%, while the predicted risk of non-Hispanic whites dropped by 1.6% and Asians dropped 4.9%.[122] The second column shows the effects of scores on the average predicted risk of the different groups using the impact of scores on predicted risk that comes from models that include controls for race, ethnicity, and income. When these score effects were used, the average predicted risk of African Americans increased by 8.9% and Hispanics by 3.5%, while the predicted risk of non-Hispanic whites decreased by 1.4% and Asians by 4.8%.[123] The change in the impact of scores on predicted risk when race, ethnicity, and income controls were included was statistically significant for all racial and ethnic groups. However, given that the use of these controls when determining the effects of scores resulted in relatively small decreases in the effect of scores on predicted risk for African Americans (10% versus

---

[122] These are the same results that were presented in Table 4.

[123] The effects of other variables are held constant between the two models. This was done by using the estimated risk effects of non-credit risk variables from the models without race, ethnicity, and income controls, and the estimated risk effects of the score deciles from the models with the controls. The estimated risk effects of the race, ethnicity, and income controls were not used to predict risk. This hybrid risk estimate produced an overall average predicted claims payout that was lower than the actual average amount of claims payouts, so every individual's predicted risk was then inflated by the ratio of actual average claims over predicted average claims.

8.9%) and Hispanics (4.2% versus 3.5%), it is apparent that most of the effect of using scores on these groups is not because scores act as a proxy for race, ethnicity, and income.

To provide a basis for comparison in evaluating the importance of these proxy effects, the FTC conducted the same analysis for several other standard risk variables. This could only be done for a small set of the risk variables in the FTC database.[124] Variables that could be used were tenure (number of years the customer has been with the company), the model year of the car, and the vehicle identification number ("VIN"), which the FTC used to obtain information on vehicle characteristics like body type and safety systems.[125] In addition, there are two risk variables in the FTC database that did not come from the company policy-level database. These are the geographic risk measure and the CLUE prior-claims data.

Table 8 shows the results of applying the FTC's proxy-effect analysis to these variables. The proxy-effect analysis was applied to these other variables in the same way it was applied to scores.[126] These other variables have much smaller effects on the average predicted risk of different racial and ethnic groups than do scores.[127] For three of the four variables, adding race, ethnicity, and income controls reduced the magnitude of the impact that the variables had on the change in predicted risk for different groups. Adding the geographic risk measure increased average predicted risk 5.4% for African Americans, 3.3% for Hispanics, and 4.4% for Asians.[128] When controls were included for race, ethnicity, and income, the impact of the geographic risk measure decreased to 4.7% for African Americans, 2.2% for Hispanics, and 3.6% for Asians. The effect of tenure on predicted risk for different groups was also reduced by adding race, ethnicity, and income controls, from 0.4% to 0.1% for African Americans, from 2.4% to 1.9% for Hispanics, and from 2.1% to 1.7% for Asians. Finally, including race, ethnicity, and income controls reduced the impact of prior claims on predicted risk from 2.4% to 2.2% for African Americans, from 0.3% to 0.2% for Hispanics, and from 1.5% to 1.4% for Asians. While these effects are small in absolute value, they are of a similar proportion to the effects that these controls have on scores' impact on the predicted risk of different racial and ethnic groups. Thus, like scores, these other risk variables also gain some predictive power from acting as proxies for race, ethnicity, or income.

---

[124] Most of the standard risk variables that came from the companies' data had large numbers of missing values, which reflects the fact that some companies did not collect or store information on some of the variables. This means that evaluating these variables is complicated by the fact that when a group of policies has "missing" as the value of a given variable, that may mean that most of the policies came from the same company. When this is true, the effects of individual variables on risk may be confounded with differences across companies in the average risk of their customers.

[125] The VINs in the FTC database were truncated, so individual cars cannot be identified. While VIN is missing for a substantial number of cars, this is mainly for cars in earlier model years. Newer model years have relatively small numbers of missing values, roughly 12%, suggesting that the missing values are unlikely to be driven primarily by differences across companies in reporting VINs.

[126] The first column of Table 8 shows the difference in predicted risk between a model that does not include the variable being tested and a model that does. The second column shows the difference between a model that does not include the variable being tested and a model that does include the variable, where the impact of the variable comes from a model that includes controls for race, ethnicity, and income.

[127] There are several reasons that could explain why the impacts of these variables on the predicted risk of different groups are not as large. It may be because the differences in these variables across groups are not (continued)

In summary, the FTC's analysis shows that credit-based insurance scores do predict risk within different racial, ethnic, and income groups. Thus, they do not act solely as a proxy for membership in these groups. Scores, however, do gain a small amount of additional predictive power because of a proxy effect. Controlling for race

---

as large as the differences for scores, because the impacts of these variables on predicted risk are not as large as the impact of scores, and/or because the impact that the inclusion of the variable has on the risk associated with other variables is not as large as the impact that scores have.

[128] Note that this is not a geographic risk measure used by any company, but rather a variable created for the purpose of this study. In addition, the geographic risk measure is not a very effective control for risk on comprehensive coverage. A better geographic risk control for comprehensive risk would likely have a larger impact on the average predicted risk of African Americans and Hispanics, for comprehensive coverage, and thus overall, given the large risk differences between African Americans and Hispanics versus non-Hispanic whites for that coverage.

and ethnicity in estimating the relationship between scores and risk causes a small reduction in the extent to which scores increase the expected risk of African Americans and Hispanics. Finally, this small proxy effect is not limited to scores, but was found for three of four other risk variables studied.

VII. ALTERNATIVE SCORING MODELS

FACTA directed the Commission to determine whether credit-based insurance scoring models could be developed that would reduce the differences in scores for consumers in protected classes relative to other consumers, yet continue to be effective predictors of risk.[129] Because race and ethnicity account for the largest differences in credit-based insurance scores among groups of consumers in the FTC database, the agency focused on constructing an effective model that decreased differences among racial and ethnic groups. To the extent practicable, the Commission also sought to build an effective model that decreased differences among income groups.

As discussed above, credit-based insurance scores are calculated using models that assign values to credit history variables to calculate numerical scores. To develop a model that effectively predicts risk while reducing differences between racial and ethnic groups, the FTC first created a baseline scoring model using the information in its database. The Commission chose variables for its baseline model with regard only to their power to generate a score that predicts risk as accurately as possible. The FTC then used a number of different techniques to try to construct alternative scoring models that were as predictive as the FTC baseline model, yet had smaller differences in scores among racial and ethnic groups.

---

[129] FACTA § 215(a)(3) (2006); 15 U.S.C. § 1681 note (2006).

The FTC was not able to develop a credit-based insurance scoring model that met the dual objectives of maintaining predictive power and decreasing the differences in scores between racial, ethnic, and income groups. This does not necessarily mean that a model could not be constructed that meets these objectives. It does strongly suggest, however, that there is no readily available scoring model or score development methodology that would do so.

A. The FTC Baseline Model

Developing a baseline model to use for comparisons is the first step in determining whether a model can produce scores that continue to predict risk but have smaller differences by race and ethnicity.[130] The FTC used claims information in its database, the non-credit risk variables in the database, and credit history variables that were appended to the insurance policy data to build the model.[131] The FTC database includes 180 credit history variables for each consumer in the development sample. This is a set of variables that ChoicePoint developed over time for its own score-building, and they are intended to capture all relevant information in a credit report.[132]

---

[130] Using either the ChoicePoint or FICO model as the base model would not be a useful test. Even a very simplistic model developed with the FTC database is likely to do better at predicting claims in the FTC database than either of those scores, because it is predicting “within sample.” That is, the model is predicting the very claims that were used to develop it.

[131] The development sample was limited to consumers for whom there is race, ethnicity, and income information in the FTC database. This demographic information was used only to develop the alternative scoring models, not the baseline model. Appendix G contains a description of the methodologies used to produce this credit-based insurance score, as well as the other scoring models discussed in this section.

[132] No ChoicePoint model uses all 180 variables, and many of these variables are not used in any model.

The Commission selected variables for its baseline model that would produce credit-based insurance scores that were effective in predicting total dollars paid out on claims per year,[133] after controlling for other non-credit risk factors, such as age and driving history. This model was constructed without giving any consideration to race, ethnicity, or income. Insurance companies and other private firms that develop scoring models likewise build their models in a "race blind" fashion.

The variables that the FTC determined produced scores that were most predictive of the claims of the consumers in its development sample are presented in Table 9.[134] It shows the fifteen variables chosen and the scoring factor assigned to each of them.[135] To calculate a score for a consumer, the factors for his or her values of each variable are multiplied together.[136]

The first five variables that enter the model each represent different aspects of a credit report: (1) Delinquencies: presence of derogatory information on the file; (2) Credit utilization: number of accounts with balance greater than 75% of the credit limit or all-time high credit balance; (3) Age of accounts: average age of bank revolving (credit card) accounts; (4) Inquiries; and (5) Type of Credit: presence of an open auto finance account in the credit report.[137] The variables that entered the model later are all variations on these categories, with the exception of a variable that measures what share of credit card accounts on the report are currently reported as "open." The category with the greatest impact on scores is delinquencies, which makes up six of the fifteen variables.[138]

---

[133] The models are intended to be predictive of claims for all major types of coverage. For this reason, claims were summed across coverages into a single measure of losses. Claims under first-party medical coverage's, "Med Pay" and personal injury protection, are also included in the "total losses" variable.

[134] The variable descriptions are proprietary and confidential information of ChoicePoint. Some variable descriptions have been made public previously. For other variables, we include only a general description of the type of variable.

[135] The Table shows the variables in the order in which they were chosen by the score-building methodology. Variables chosen earlier are generally those that provide greater predictive power to the scoring model.

[136] The resulting score is the inverse of the relative predicted risk for the consumer. The inverse is used so that higher scores are associated with lower predicted risk.

[137] An auto finance account is an account with a lender associated with a car company, like GMAC or Ford Motor Credit.

The scores the FTC baseline model produces did predict risk. Figure 17 shows the relationship between total claims paid out and the FTC credit-based insurance score for the four major types of automobile coverage. Each graph shows three lines: (1) the average total amount paid on claims by score decile in the development sample; (2) the estimated relationship between scores and claims in the development sample from models controlling for other risk factors; and, (3) the average total amount paid on claims by score decile in CLUE data for the period June 2001 to December 2001, for people who were not in the development sample (an "out of sample" check).[139] If the model generated scores that effectively predicted risk, then the lines on the graph should slope downward to the right. The FTC baseline model produced results consistent with this expected pattern. For example, for bodily injury liability coverage, consumers in the lowest 10% of scores were more than three and a half times riskier than consumers in the highest 10% of scores. Even for property damage liability claims, which have the weakest relationship with the FTC score, consumers in the lowest 10% of scores of the development sample were more than twice as risky as consumers in the highest 10% of scores. Consequently, the FTC baseline model is an effective predictor of risk. Figure 17 also shows that the FTC baseline model predicts risk for people outside the development sample. This result is important in that it shows the FTC baseline model scores do not simply predict the claims that were used to develop the model.

---

[138] In looking over the model, it is important to keep in mind that a piece of information in a credit report can be represented in multiple ways and affect multiple variables. This means that care must be taken when interpreting some of the results. For example, the score factors for variable C show that delinquencies on a particular kind of account actually lead to a better score, which seems very strange in isolation. But, it simply means that, in these data, a delinquency on that type of account is less indicative of risk than delinquencies on other kinds of accounts, since there is another variable in the model that is a broad measure of delinquencies and has a large negative impact on score.

[139] The development sample consists only of the sub-sample of the FTC database for which we obtained SSA race and ethnicity data. The development sub-sample includes everyone who had a claim in the company data, so there was no way to use the company data to look at claims outside of the development sample. Instead, we use CLUE data on claims for a different time period. We were able to use data on roughly 800,000 policies for these checks.

To establish a baseline for evaluating the results of other models, the FTC also measured the extent to which its model resulted in differences in scores among racial and ethnic groups. Figure 18 shows how the four groups were distributed across the range of FTC baseline-model scores. The horizontal axis shows score deciles, and the vertical axis shows the share of each group that fell in each decile. The deciles were defined using the overall distribution of scores, so if a given group had the same distribution of scores as the overall sample, 10% of that group's population would fall in each decile. Figure 18 shows that the FTC baseline model produced lower scores for African Americans and Hispanics than for non-Hispanic whites and Asians.[140]

Table 9 also shows the breakdown of the different racial and ethnic groups across the variables used in the FTC's baseline model. The variables that show large differences across racial and ethnic group are those relating to payment history (*e.g.*, delinquencies) and public records, and the variable for the share of accounts with high balances relative to credit limits. The inclusion of these variables in the FTC baseline model explains why African Americans and Hispanics had lower scores than non-Hispanic whites and Asians.

---

[140] Note that these differences across racial and ethnic groups for the FTC baseline model are very similar to those for the ChoicePoint scores discussed above, with the only substantial difference being that Asians were less well represented in the higher score categories for the FTC baseline model than for ChoicePoint scores.

B. Alternative Scoring Models

1. “Race Neutral” Scoring Models

The FTC credit-based insurance scoring model described in the previous section provides a baseline for evaluating alternative models. To construct a model that was “neutral” with respect to race, ethnicity, and income, the FTC created a model in which it controlled directly for these factors.[141] “Neutral” in this context means that while the scores produced by the model still may vary across groups, the variables used in the scoring model should not derive predictive power from a relationship with race, ethnicity, or income. Controls mitigate the impact of credit history variables that differ widely among different racial, ethnic, or income groups, if those variables derive a substantial portion of their power to predict losses from those differences. If controls are used for race, ethnicity, and income, these variables become less predictive of risk. With this loss in predictive power, these variables are either not selected for a scoring model at all, or, if selected, they are not given as much weight.

Table 10 shows the scoring model that was produced if controls for race, ethnicity, and income were used in the model building process. Most significantly, the variables selected in this model that controls for race (a race “neutral” model) are extremely similar to those in the FTC baseline model (a race “blind” model). Specifically, only two of the fifteen variables are different between these two models, and these two particular variables have a relatively weak effect on predicting risk. Despite controlling for race, ethnicity, and income, a very similar set of credit history variables thus were found to be most predictive of claims. Even though some of these variables have large differences across racial and ethnic groups, the variables were chosen not because they vary by race, ethnicity, or income.

---

[141] Several authors have proposed this approach. *See* Elaine Fortowsky and Michael Lacour-Little, *Credit Scoring and Disparate Impact* (Dec. 31, 2001), *available at* http://fic.wharton.upenn.edu/fic/lacourpaper.pdf; Stephen L. Ross and John Yinger, THE COLOR OF CREDIT: MORTGAGE DISCRIMINATION, RESEARCH METHODOLOGY, AND FAIR LENDING ENFORCEMENT (2002).

The Commission tried another approach to developing a race "neutral" model to compare to the FTC baseline model. We constructed a credit-based insurance scoring model using a development sample that included only non-Hispanic whites. Because there were no other racial or ethnic groups in the sample used to construct such a model, the predictive power of the variables selected cannot be attributed to any relationship with race or ethnicity.

Table 11 shows the variables selected when a model was built using only non-Hispanic whites as the development sample. Upon first examination, the variables selected for this model appear quite different from the variables in the FTC baseline model. Eight of the fifteen variables are different, including the variable with the second greatest impact. However, there is an important similarity between the variables in these two models. The same *types* of variables were found to be the most important: delinquencies, inquiries, measures of high debt burden, age of the credit report, and type of credit.

Both race-neutral models that the FTC developed predict risk within the development sample about as well as the FTC baseline model. Figure 19 compares the results for each of these models for each of the four types of automobile insurance coverage.[142] These graphs show that the FTC baseline model (a race blind model) produced very similar results for each type of coverage as models that controlled for race and ethnicity or that were developed using only non-Hispanic whites (race neutral models). Given the similarity between the types of variables selected for use in these models, it is not surprising that these scores have comparable power in predicting risk.

---

[142] Although only non-Hispanic whites were used to develop the "non-Hispanic whites" model, the results shown here are for the complete development sample.

Just as their risk prediction is comparable to that of the FTC baseline model, the race neutral models also display large differences in scores among racial and ethnic groups. Figure 20 shows the distribution of scores for the different racial and ethnic groups for the two race neutral models and the FTC baseline model. To facilitate comparisons, each graph shows the results for all three models for a single racial or ethnic group. For all groups except Asians, the distribution of people across deciles was nearly identical for the three scoring models. For Asians, the FTC baseline model and the model developed using controls for race, ethnicity, and income gave very similar results. The model built using only non-Hispanic whites, however, produced a distribution of scores for Asians that was more skewed towards lower scores.

In short, these comparisons show that, although the race neutral models that the FTC built accurately predict risk, they do not decrease the differences in credit-based insurance scores among racial and ethnic groups.

2. Model Discounting Variables with Large Differences by Race and Ethnicity

In addition to developing race neutral models as possible alternatives, the FTC also constructed alternative models that tried more directly to avoid selecting variables with large differences among racial and ethnic groups. In building such models, the FTC measured not just how well a given variable predicted claims, but how well it predicted race and ethnicity. The FTC then chose the variables that contributed the most to predicting risk and the least to predicting race and ethnicity.

Table 12 shows one of the models developed using this approach. It is very different from the models described in the previous two sections. Most significantly, there are no variables that relate directly to delinquencies, which Tables 9 – 11 showed varied a great deal among racial and ethnic groups. Most variables selected relate to the number and type of accounts that a consumer has. In addition, the discounted model includes variables that relate to the age of the credit account and total indebtedness.

Figure 21 shows that the discounted model is much less predictive of risk than the FTC baseline model for each of the four types of automobile insurance coverage. The discounted model does produce credit-based insurance scores that predict risk. However, each of these graphs shows that the relationship between the credit score and risk is much weaker (flatter) for the discounted model than for the FTC baseline model. This shows that this process of avoiding variables with large differences between groups resulted in a model that is substantially less effective as a predictor of risk than the FTC baseline model.

Figure 22 compares scores for each racial and ethnic group based on the results obtained from the discounted model and the FTC baseline model. The model that assigns consumers in a racial or ethnic group most closely to 10% in each decile (*i.e.*, a flat line at 10% on the vertical axis) shows the least differences based on race and ethnicity. Each of these graphs shows that the discounted model resulted in scores with smaller differences between members of racial and ethnic minority groups than did the FTC baseline model. These differences were most substantial for African Americans. While they were still slightly over-represented in the lower score categories, the scores from the discounted model showed 14% of African Americans are in the bottom 10% of scores. The scores from the FTC baseline model, in contrast, showed 27% of African Americans in the bottom 10% of scores. Although the discounted model did substantially reduce the differences in scores among members of racial and ethnic groups, as discussed above, it also provides far less effective risk prediction.

In summary, the FTC's inability to build a model that produces scores that continues to predict risk accurately at the same time as narrowing the differences in scores among racial and ethnic minority groups are by no means definitive. Perhaps someone could develop a model that meets both of these objectives. The FTC's inability to build to such a model, however, strongly suggests that there is no readily available approach for doing so.

VIII. CONCLUSION

The FTC's analysis demonstrates that credit-based insurance scores are effective predictors of risk under automobile insurance policies. Using scores is likely to make the price of insurance conform more closely to the risk of loss that consumers pose, resulting, on average, in higher-risk consumers paying higher premiums and lower-risk consumers paying lower premiums. It has not been clearly established why scores are predictive of risk.

Credit-based insurance scores may benefit consumers overall. Scores may permit insurance companies to evaluate risk with greater accuracy, which may make them more willing to offer insurance to higher-risk consumers. Scores also may make the process of granting and pricing insurance quicker and cheaper, cost savings that may be passed on to consumers in the form of lower premiums. However, little hard data was submitted or available to the FTC to quantify the magnitude of these potential benefits to consumers.

Credit-based insurance scores are distributed differently among racial and ethnic groups. The FTC's analysis revealed that the use of scores for consumers whose information was included in the FTC's database caused the average predicted risk for African Americans and Hispanics to increase by 10% and 4.2%, respectively, while it caused the average predicted risk for non-Hispanic whites and Asians to decrease by 1.6% and 4.9%, respectively. These changes in predicted risk are likely to have an effect on the insurance premiums that these groups on average pay.

Credit-based insurance scores predict risk within racial, ethnic, and income groups. Scores have only a small effect as a "proxy" for membership in racial and ethnic groups in estimating of insurance risk, remaining strong predictors of risk when controls for race, ethnicity and income are included in risk models. The FTC's analysis revealed that the use of scores for consumers whose information was included in the FTC's database caused the average predicted risk for African Americans and Hispanics to increase by 10% and 4.2%, respectively. The Commission's analysis also showed that using the effects of scores on predicted risk that come from models that include controls for race, ethnicity, and income caused scores to increase the average predicted risk for African Americans and Hispanics by 8.9% and 3.5%, respectively. The difference between these two predictions for these two groups (1.1% and 0.7%, respectively) shows that a relatively small portion of the impact of scores on these groups comes from scores acting as a proxy for race, ethnicity, and income.

Finally, the FTC was not able to develop an alternative credit-based insurance scoring model that would continue to predict risk effectively, yet decrease the differences in scores on average among racial and ethnic groups. This does not mean that a model could not be constructed that meets both of these objectives. It does strongly suggest, however, that there is no readily available scoring model that would do so.

TABLES

TABLE 1.
Typical Information Used in Credit-Based Insurance Scoring Models

Performance on Credit Obligations

- Late payments/Delinquencies (-)
- Collections (generally non-medical) (-)
- Public records (judgments or bankruptcies) (-)

Credit-Seeking Behavior

- Inquiries (generally non-insurance, non-medical) (-)
- New accounts (-)

Use of Credit

- Ratio of outstanding balances to available credit (-)

Length of Credit History

- Age of oldest account (+)
- Average age of all accounts (+)

Types of Credit Used

- Department store trade lines (-)
- Oil Company trade lines (-)
- Travel and Entertainment trade lines (-)
- Share of trade lines that are major bank credit cards or mortgages (+)

Note: (-) indicates that high values typically lead to a riskier score, and the converse for (+).

TABLE 2.
Claim Frequency, Claim Severity, and Average Total Amount Paid on Claims

Score Decile	Average Number of Claims Per Year of Coverage (per hundred) (a)	Average Cost per Claim (b)	Average Total Paid on Claims Per Year of Coverage [(a) x (b)] (c)
Property Damage Liability Coverage			
1	5.65	$2,100	$119
2	4.86	2,119	103
3	4.51	2,105	95
4	4.21	2,078	88
5	4.09	1,982	81
6	3.85	2,028	78
7	3.55	2,006	71
8	3.34	1,994	67
9	3.40	2,062	70
10	3.17	1,981	63
Overall	*4.06*	*$2,053*	*$83*
Bodily Injury Liablility Coverage			
1	1.79	$8,560	$153
2	1.59	10,002	159
3	1.39	7,798	109
4	1.39	7,993	111
5	1.19	7,940	95
6	1.01	8,892	89
7	0.91	8,538	78
8	0.89	8,760	78
9	0.85	9,127	78
10	0.77	8,372	64
Overall	*1.18*	*$8,609*	*$101*

(continued…)

TABLE 2.
Claim Frequency, Claim Severity, and Average Total Amount Paid on Claims (Continued)

	Score Decile	Average Number of Claims Per Year of Coverage (per hundred) (a)	Average Cost per Claim (b)	Average Total Paid on Claims Per Year of Coverage [(a) x (b)] (c)
Collision Coverage				
	1	11.80	$2,364	$279
	2	9.53	2,201	210
	3	8.57	2,174	186
	4	8.09	2,060	167
	5	7.45	2,014	150
	6	6.86	2,057	141
	7	6.47	2,006	130
	8	6.18	1,965	122
	9	6.11	2,003	122
	10	5.38	2,004	108
	Overall	*7.64*	*$2,112*	*$161*
Comprehensive Coverage				
	1	11.50	$1,032	$119
	2	9.69	879	85
	3	9.06	828	75
	4	9.06	773	70
	5	8.34	773	64
	6	8.07	752	61
	7	7.46	774	58
	8	7.42	718	53
	9	7.03	722	51
	10	6.95	688	48
	Overall	*8.44*	*$807*	*$68*

Note: All numbers on this table represent actual means (*i.e.* , not derived from any risk modelling procedure).

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 3.
Median Income and Age, and Gender Make-Up, by Race and Ethnicity

	Median Tract Income (a)	Median Age (b)	Percent Male (c)
African Americans	$34,876	46	48%
Hispanics	$38,475	42	60%
Asians	$50,953	42	72%
Non-Hispanic Whites	$44,356	48	68%

Note: Age and gender are measured at the individual level. See section VI.A.2 of the report for a discussion of how the age of the individual was determined. Neighborhood income is the median for the Census tract where the individual lives. See Appendix C for details on the data sources and the construction of the database.

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 4.
Change in Predicted Amount Paid on Claims from Using Credit-Based Insurance Scores, by Race and Ethnicity

	Share With a Decrease (a)	Share With an Increase (b)	Percent Change in Mean Predicted Risk (c)
African Americans	36%	64%	10.0%
Hispanics	47%	53%	4.2%
Asians	66%	34%	- 4.9%
Non-Hispanic Whites	62%	38%	- 1.6%
Overall	*59%*	*41%*	*0.0%*

Note: Predicted change in the amount paid on claims was estimated by comparing individuals' predicted total claims from risk models that include ChoicePoint Attract Standard Auto credit-based insurance scores with risk models that do not include scores. (By construction, the average of all changes for the sample is zero.) Both of these models include a standard set of risk variables as controls, and were run separately for property damage liability, bodily injury liability, collision, and comprehensive coverage. In the final step we sum the predicted dollar risks for all four types of insurance coverage with and without the use of credit-based insurance scores. See section VI.A.3 of the report for additional details on this analysis. Modeling details and a description of the variables included in the models are provided in Appendix D.

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 5.
Estimated Relative Amount Paid on Claims, by Race, Ethnicity, and Neighborhood Income

	Property Damage Liability Coverage (a)	Bodily Injury Liability Coverage (b)	Collision Coverage (c)	Comprehensive Coverage (d)
Race and Ethnicity				
African Americans	1.01	1.48 *	1.43 *	1.63 *
Hispanics	1.11	1.25 *	1.33 *	1.45 *
Asians	1.17 *	1.11	1.30 *	0.96
Non-Hispanic Whites	1.00	1.00	1.00	1.00
Neighborhood Income				
Low	0.97	1.01	1.05	1.16 *
Middle	0.95 *	1.02	0.99	1.06 *
High	1.00	1.00	1.00	1.00

Asterisks indicate statistically significantly different from base category at 5% level.

Notes:

1) For each variable – *i.e.* race and ethnicity, and neighborhood income – estimated amount paid on claims per year of coverage is measured relative to a base category. For race and ethnicity, the base category is non-Hispanic whites; and, for neighborhood income the base category is “high income” neighborhood.

2) Estimated relative amounts paid out on claims per year of coverage for each race, ethnicity and neighborhood income category in each column are derived from Tweedie GLMs (Generalized Linear Models); which here include a set of standard risk variables as controls, but not score. Since our GLM models are multiplicative, the relativities shown on this table are equivalent to the exponentiated regression coefficients of the indicator variables for these categories. Modeling details and a description of the variables included in the models are provided in Appendix D.

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 6.
Estimated Relative Amount Paid on Claims, by Score Decile, Race, Ethnicity, and Neighborhood Income

	Property Damage Liability Coverage		Bodily Injury Liability Coverage	
	(a)	(b)	(c)	(d)
Score Decile				
1	1.70 *	1.73 *	2.20 *	2.10 *
2	1.52 *	1.53 *	2.14 *	2.07 *
3	1.43 *	1.44 *	1.75 *	1.72 *
4	1.35 *	1.35 *	1.66 *	1.65 *
5	1.24 *	1.24 *	1.37 *	1.36 *
6	1.23 *	1.23 *	1.26 *	1.26 *
7	1.13 *	1.12 *	1.15	1.14
8	1.07	1.07	1.13	1.13
9	1.12 *	1.12 *	1.21	1.21
10	1.00	1.00	1.00	1.00
Race and Ethnicity				
African Americans	-	0.93	-	1.29 *
Hispanics	-	1.06	-	1.15
Asians	-	1.20 *	-	1.15
Non-Hispanic Whites	-	1.00	-	1.00
Neighborhood Income				
Low	-	0.96	-	0.98
Middle	-	0.94 *	-	1.00
High	-	1.00	-	1.00

Asterisks indicate statistically significantly different from base category at 5% level.

Coefficients in dashed boxes are statistically significantly different across models (within a given coverage type) at the 5% level.

(continued. . .)

TABLE 6.
Estimated Relative Amount Paid on Claims, by Score Decile, Race, Ethnicity, and Neighborhood Income (Continued)

	Collision Coverage		Comprehensive Coverage	
	(e)	(f)	(g)	(h)
Score Decile				
1	2.03 *	1.93 *	1.95 *	1.74 *
2	1.65 *	1.59 *	1.43 *	1.33 *
3	1.52 *	1.48 *	1.33 *	1.26 *
4	1.39 *	1.36 *	1.28 *	1.23 *
5	1.27 *	1.25 *	1.19 *	1.16 *
6	1.26 *	1.25 *	1.15 *	1.12 *
7	1.16 *	1.15 *	1.12 *	1.10 *
8	1.09	1.08	1.05	1.04
9	1.12 *	1.12 *	1.01	0.99
10	1.00	1.00	1.00	1.00
Race and Ethnicity				
African Americans	-	1.26 *	-	1.46 *
Hispanics	-	1.24 *	-	1.36 *
Asians	-	1.33 *	-	0.97
Non-Hispanic Whites	-	1.00	-	1.00
Neighborhood Income				
Low	-	1.01	-	1.13 *
Middle	-	0.97	-	1.04
High	-	1.00	-	1.00

Asterisks indicate statistically significantly different from base category at 5% level.

Coefficients in dashed boxes are statistically significantly different across models (within a given coverage type) at the 5% level.

Notes:

1) For each variable – score, race and ethnicity, and neighborhood income – estimated amount paid on claims per year of coverage is measured relative to a base category. For scores, the base category is the 10th (highest) decile of scores; for race and ethnicity, the base category is non-Hispanic whites; and, for neighborhood income the base category is “high income” neighborhood.

2) Estimated relative amounts paid out on claims per year of coverage for each race, ethnicity and neighborhood income category in each column are derived from Tweedie GLMs (Generalized Linear Models); which here include a set of standard risk variables as controls, as well as score deciles. Since our GLM models are multiplicative, the relativities shown on this table are equivalent to the exponentiated regression coefficients of the indicator variables for these categories. Modeling details and a description of the variables included in the models are provided in Appendix D.

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 7.
Change in Predicted Amount Paid on Claims from Using Credit-Based Insurance Scores Without and With Controls for Race, Ethnicity, and Income, by Race and Ethnicity

	Average Score Effect From Model Without Race, Ethnicity, and Income Controls (a)	Average Score Effect from Model With Race, Ethnicity, and Income Controls (b)
African Americans	10.0%	8.9%
Hispanics	4.2%	3.5%
Asians	- 4.9%	-4.8%
Non-Hispanic Whites	- 1.6%	-1.4%

Numbers for all race and ethnicity groups are statistically significantly different across the models in columns (a) and (b) at the 5% level.

Notes:

Column (a): Results in this column come from the same analysis that was used to create Table 4. Predicted change in the amount paid on claims was estimated by comparing individual predicted risk from risk models that include ChoicePoint Attract Standard Auto credit-based insurance scores with risk models that do not include scores. All models include a standard set of risk variables as controls, and were run separately for property damage liability, bodily injury liability, collision, and comprehensive coverage (in the final step we sum the predicted dollar risks for all four types of insurance coverage); the same is true for column (b). This procedure is described in section VI.A.3 of the report. Modeling details and a description of the variables included in the models are provided in Appendix D.

Column (b): Results in this column are calculated by combining the estimated risk effects of the score deciles from models with controls for race, ethnicity, and income with the estimated risk effects of non-credit risk variables from the models used in column (a), which do not include these additional controls. The estimated risk effects of race, ethnicity, and income were not used to predict risk. This hybrid risk estimate produced an overall average predicted claims payout that was lower than the actual sample average amount of claims payouts, so every individual’s predicted risk was then inflated by the ratio of actual average claims over predicted average claims.

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 8.
Change in Predicted Amount Paid on Claims from Using Other Risk Variables, Without and With Controls for Race, Ethnicity, and Income, by Race and Ethnicity

	Average Effect of Variable Without Race, Ethnicity, and Income Controls (a)	Average Effect of Variable With Race, Ethnicity, and Income Controls (b)
Geographic Risk		
African Americans	5.4%	4.7%
Hispanics	3.3%	2.2%
Asians	4.4%	3.6%
Non-Hispanic Whites	-1.3%	-1.0%
Tenure		
African Americans	0.4%	0.1%
Hispanics	2.4%	1.9%
Asians	2.1%	1.7%
Non-Hispanic Whites	-0.5%	-0.4%
Prior Claims		
African Americans	2.4%	2.2%
Hispanics	0.3%	0.2%
Asians	1.5%	1.4%
Non-Hispanic Whites	-0.3%	-0.3%

(continued…)

TABLE 8.
Change in Predicted Amount Paid on Claims from Using Other Risk Variables, Without and With Controls for Race, Ethnicity, and Income, by Race and Ethnicity (Continued)

	Average Effect of Variable Without Race, Ethnicity, and Income Controls (a)	Average Effect of Variable With Race, Ethnicity, and Income Controls (b)
Model Year & Other Car Attributes		
African Americans	-1.0%	-1.2%
Hispanics	0.5%	0.5%
Asians	2.8%	2.6%
Non-Hispanic Whites	0.0%	0.0%

Notes:

Column (a): Results in this column come from an analysis similar to that used to create Table 4 for score. Predicted change in the amount paid on claims was estimated by comparing individual predicted risk from risk models that included the particular variable being analyzed here with risk models that did not include the variable. All models include the standard set of risk controls (including score), and were run separately for property damage liability, bodily injury liability, collision, and comprehensive coverage (in the final step we sum the predicted dollar risks for all four types of insurance coverage); the same is true for column (b). This procedure is described in section VI.A.3 of the report. Modeling details and a description of the variables included in the models are provided in Appendix D.

Column (b): Results in this column are calculated by combining the estimated risk effects of the variable being analyzed from models with controls for race, ethnicity, and income with the estimated risk effects of all other risk variables from the models used in column (a), which do not include these additional controls. The estimated risk effects of race, ethnicity, and income were not used to predict risk. This hybrid risk estimate produced an overall average predicted claims payout that was lower tha the actual sample average amount of claims payouts, so every individual’s predicted risk was then inflated by the ratio of actual average claims over predicted average claims.

Source: Analysis of FTC Automobile Insurance Policy Database

TABLE 9.
Baseline Credit-Based Insurance Scoring Model Developed by the FTC

1) *Variable A: Presence of Certain Delinquencies or Adverse Public Records on the Credit File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.14	84.5%	56.0%	69.9%	83.0%
1 or more	1.00	15.5%	44.0%	30.1%	17.0%

2) Number of Accounts with Balance Greater than 75% of High Credit (Credit Limit)

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.25	43.2%	20.3%	28.9%	43.0%
1 - 2	1.16	24.9%	21.8%	24.6%	24.4%
2 - 3	1.09	13.1%	17.3%	15.9%	14.3%
3 - 6	1.04	14.0%	27.6%	23.3%	13.2%
6 or More	1.00	4.8%	13.0%	7.4%	5.1%

3) Average Number of Months Bank Revolving Accounts Have Been Open

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.16	3.2%	5.6%	5.5%	2.7%
0 - 24	0.67	3.6%	9.2%	10.0%	6.3%
24 - 51	0.80	10.4%	18.6%	18.6%	16.1%
51 - 64	0.83	9.4%	10.3%	12.0%	11.5%
64 - 99	0.84	34.8%	27.8%	31.3%	36.8%
99 - 205	0.87	36.2%	26.3%	21.5%	25.4%
205 or More	1.00	2.4%	2.1%	1.2%	1.3%

(continued. . .)

TABLE 9.
Baseline Credit-Based Insurance Scoring Model Developed by the FTC (Continued)

4) *Variable B: Relates to the Number of Inquiries on the File*

Category	Factor	Share in each category, by race or ethnicity Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade lines	1.30	34.1%	21.6%	15.9%	19.8%
0	1.31	16.6%	13.9%	14.4%	16.2%
1 - 2	1.29	22.0%	22.8%	20.0%	21.3%
2 - 4	1.20	17.8%	23.2%	25.9%	23.6%
4 - 7	1.13	7.3%	12.6%	16.0%	12.7%
7 or more	1.00	2.4%	5.9%	7.8%	6.4%

5) Number of Open Auto Finance Accounts

Category	Factor	Share in each category, by race or ethnicity Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.13	90.0%	84.1%	88.0%	85.3%
0 or more	1.00	10.0%	15.9%	12.0%	14.7%

6) Number of Accounts 30 Days Late or Worse in the Last 12 Months

Category	Factor	Share in each category, by race or ethnicity Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.36	77.1%	47.4%	61.4%	75.8%
1 - 9	1.13	22.2%	50.3%	37.4%	23.6%
10 or more	1.00	0.7%	2.2%	1.2%	0.7%

7) *Variable C: Presence of Delinquencies on a Particular Kind of Account*

Category	Factor	Share in each category, by race or ethnicity Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.99	25.1%	23.8%	20.0%	25.1%
0	0.78	71.8%	67.0%	72.5%	71.4%
1 or more	1.00	3.1%	9.2%	7.5%	3.5%

(continued. . .)

TABLE 9.
Baseline Credit-Based Insurance Scoring Model Developed by the FTC (Continued)

8) Number of Department Store Accounts

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.00	25.1%	23.8%	20.0%	25.1%
1 or more	1.17	72.3%	71.2%	74.4%	71.7%
6 or more	1.00	2.5%	5.0%	5.6%	3.2%

9) Share of all Bank Revolving Accounts that are Open

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.75	5.4%	9.8%	8.9%	4.8%
0 - .135	0.89	2.8%	4.3%	3.5%	2.8%
> .135	1.00	91.7%	85.8%	87.7%	92.3%

10) *Variable D: Presence of a Particular Kind of Delinquency on the Account*

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.16	83.9%	54.1%	68.9%	82.3%
1 or more	1.00	16.1%	45.9%	31.1%	17.7%

11) Age of Youngest Account (Months)

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 6	0.82	31.7%	37.2%	38.3%	37.4%
6 - 9	0.87	15.1%	18.0%	18.0%	15.6%
9 - 20	0.89	26.3%	26.7%	25.0%	26.0%
20 or more	1.00	26.9%	18.1%	18.7%	21.0%

(continued. . .)

TABLE 9.
Baseline Credit-Based Insurance Scoring Model Developed by the FTC (Continued)

12) *Variable E: Relates to the Number of Accounts in the Credit File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.89	0.2%	0.4%	0.5%	0.0%
0 - 3	0.82	3.7%	5.6%	4.6%	3.7%
3 or more	1.00	96.2%	94.1%	94.9%	96.3%

13) *Variable F: A Ratio Relating to Delinquencies*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - .02	1.17	94.6%	80.7%	88.4%	93.8%
.02 - .14	1.20	2.7%	9.4%	6.0%	2.8%
> .14	1.00	2.8%	9.9%	5.6%	3.5%

14) Number of Bank Revolving Accounts Ever Bad Debt

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.62	3.0%	5.0%	5.2%	2.4%
0	0.90	90.1%	75.3%	81.8%	89.9%
1 or more	1.00	6.9%	19.8%	12.9%	7.7%

15) Number of Open Oil Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.92	91.6%	93.7%	88.7%	91.2%
0 or more	1.00	8.4%	6.3%	11.3%	8.8%

Notes:
1) Variables in italics have not been described publicly, and ChoicePoint considers the descriptions of those variables to be proprietary information.

2) This scoring model was developed to use credit history information to predict the relative risk posed by individuals, where risk is defined as expected total dollars that would be paid out on claims in a year. To calculate a score for a given individual with this model, the appropriate factors for each of the 15 variables are multiplied together. The resulting product is the inverse of the estimated relative riskiness of the individual, based on the individual's credit history. See Appendix E for a detailed discussion of the score-building process.

TABLE 10.
Credit-Based Insurance Scoring Model Developed by the FTC by Including Controls for Race, Ethnicity, and Neighborhood Income in the Score-Building Process

1) *Variable A: Presence of Certain Delinquencies or Adverse Public Records on the Credit File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.14	84.5%	56.0%	69.9%	83.0%
1 or more	1.00	15.5%	44.0%	30.1%	17.0%

2) Number of Accounts with Balance Greater than 75% of High Credit (Credit Limit)

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.26	43.2%	20.3%	28.9%	43.0%
1 - 2	1.17	24.9%	21.8%	24.6%	24.4%
2 - 3	1.11	13.1%	17.3%	15.9%	14.3%
3 - 6	1.05	14.0%	27.6%	23.3%	13.2%
6 or More	1.00	4.8%	13.0%	7.4%	5.1%

3) Average Number of Months Bank Revolving Accounts Have Been Open

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.13	3.2%	5.6%	5.5%	2.7%
0 - 24	0.67	3.6%	9.2%	10.0%	6.3%
24 - 51	0.80	10.4%	18.6%	18.6%	16.1%
51 - 64	0.82	9.4%	10.3%	12.0%	11.5%
64 - 99	0.84	34.8%	27.8%	31.3%	36.8%
99 - 205	0.87	36.2%	26.3%	21.5%	25.4%
205 or More	1.00	2.4%	2.1%	1.2%	1.3%

4) Number of Open Auto Finance Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.12	90.0%	84.1%	88.0%	85.3%
0 or more	1.00	10.0%	15.9%	12.0%	14.7%

(continued. . .)

TABLE 10.
Credit-Based Insurance Scoring Model Developed by the FTC by Including Controls for Race, Ethnicity, and Neighborhood Income in the Score-Building Process (Continued)

5) *Variable B: Relates to the Number of Inquiries on the File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade lines	1.30	34.1%	21.6%	15.9%	19.8%
0	1.31	16.6%	13.9%	14.4%	16.2%
1 - 2	1.28	22.0%	22.8%	20.0%	21.3%
2 - 4	1.20	17.8%	23.2%	25.9%	23.6%
4 - 7	1.13	7.3%	12.6%	16.0%	12.7%
7 or more	1.00	2.4%	5.9%	7.8%	6.4%

6) Number of Accounts 30 Days Late or Worse in the Last 12 Months

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.35	77.1%	47.4%	61.4%	75.8%
1 - 9	1.14	22.2%	50.3%	37.4%	23.6%
10 or more	1.00	0.7%	2.2%	1.2%	0.7%

7) *Variable C: Presence of Delinquencies on a Particular Kind of Account*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.97	25.1%	23.8%	20.0%	25.1%
0	0.78	71.8%	67.0%	72.5%	71.4%
1 or more	1.00	3.1%	9.2%	7.5%	3.5%

8) Share of all Bank Revolving Accounts that are Open

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.76	5.4%	9.8%	8.9%	4.8%
0 - .135	0.89	2.8%	4.3%	3.5%	2.8%
> .135	1.00	91.7%	85.8%	87.7%	92.3%

(continued. . .)

TABLE 10.
Credit-Based Insurance Scoring Model Developed by the FTC by Including Controls for Race, Ethnicity, and Neighborhood Income in the Score-Building Process (Continued)

9) Number of Department Store Accounts

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.00	25.1%	23.8%	20.0%	25.1%
1 or more	1.15	72.3%	71.2%	74.4%	71.7%
6 or more	1.00	2.5%	5.0%	5.6%	3.2%

10) Age of Youngest Account (Months)

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 6	0.81	31.7%	37.2%	38.3%	37.4%
6 - 9	0.87	15.1%	18.0%	18.0%	15.6%
9 - 20	0.89	26.3%	26.7%	25.0%	26.0%
20 or more	1.00	26.9%	18.1%	18.7%	21.0%

11) *Variable G: Relates to the Number of Accounts in the Credit File*

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 2	0.89	5.0%	9.9%	6.6%	4.1%
2 or more	1.00	95.0%	90.1%	93.4%	95.9%

12) *Variable D: Presence of a Particular Kind of Delinquency on the Account*

Category	Factor	Share in each category, by race or ethnicity: Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.16	83.9%	54.1%	68.9%	82.3%
1 or more	1.00	16.1%	45.9%	31.1%	17.7%

(continued. . .)

TABLE 10.

Credit-Based Insurance Scoring Model Developed by the FTC by Including Controls for Race, Ethnicity, and Neighborhood Income in the Score-Building Process (Continued)

13) Number of Open Personal Finance Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.90	82.0%	66.4%	73.6%	82.9%
0 - 2	0.97	14.9%	24.6%	21.5%	14.2%
2 or more	1.00	3.1%	9.0%	4.9%	2.9%

14) *Variable F: A Ratio Relating to Delinquencies*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - .02	1.17	94.6%	80.7%	88.4%	93.8%
.02 - .14	1.20	2.7%	9.4%	6.0%	2.8%
> .14	1.00	2.8%	9.9%	5.6%	3.5%

15) Number of Bank Revolving Accounts Ever Bad Debt

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.62	3.0%	5.0%	5.2%	2.4%
0	0.90	90.1%	75.3%	81.8%	89.9%
1 or more	1.00	6.9%	19.8%	12.9%	7.7%

Notes:
1) Variables in italics have not been described publicly, and ChoicePoint considers the descriptions of those variables to be proprietary information.

2) This scoring model was developed to use credit history information to predict the relative risk posed by individuals, where risk is defined as expected total dollars that would be paid out on claims in a year. To calculate a score for a given individual with this model, the appropriate factors for each of the 15 variables are multiplied together. The resulting product is the inverse of the estimated relative riskiness of the individual, based on the individual's credit history. This scoring model was developed by including controls for race, ethnicity, and neighborhood income during the process of selecting variables for the scoring model, and when estimating the final factors that are applied to the credit history variables. See Appendix E for a detailed discussion of the score-building process.

TABLE 11.
Credit-Based Insurance Scoring Model Developed by the FTC Using a Sample of Only Non-Hispanic White Insurance Customers

1) *Variable A: Presence of Certain Delinquencies or Adverse Public Records on the Credit File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.23	84.5%	56.0%	69.9%	83.0%
1 or more	1.00	15.5%	44.0%	30.1%	17.0%

2) *Variable B: Relates to the Number of Inquiries on the File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade lines	1.25	34.1%	21.6%	15.9%	19.8%
0 - 2	1.25	38.5%	36.7%	34.3%	37.5%
2 or more	1.14	21.5%	29.5%	32.0%	29.3%
5 or more	1.00	6.0%	12.3%	17.8%	13.4%

3) Total Average Debt Burden

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
Invalid past due amount	0.89	0.6%	0.8%	0.9%	0.7%
0 - .19	1.20	41.7%	18.4%	26.2%	44.0%
.19 - .46	1.13	25.5%	22.5%	25.5%	24.7%
.46 - .81	1.06	24.4%	38.8%	33.8%	23.7%
> .81	1.00	7.7%	19.4%	13.6%	6.8%

4) Age of Youngest Account (Months)

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 6	0.84	31.7%	37.2%	38.3%	37.4%
6 - 14	0.90	30.3%	35.5%	34.4%	31.1%
14 or more	1.00	38.0%	27.3%	27.3%	31.4%

(continued. . .)

TABLE 11.
Credit-Based Insurance Scoring Model Developed by the FTC Using a Sample of Only Non-Hispanic White Insurance Customers (Continued)

5) Number of Accounts 30 Days Late in the Last 24 Months

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.15	83.9%	65.3%	73.9%	83.7%
1 or more	1.00	16.1%	34.7%	26.1%	16.3%

6) Share of all Bank Revolving Accounts that are Open

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.82	5.4%	9.8%	8.9%	4.8%
0 or more	1.00	94.6%	90.2%	91.1%	95.2%

7) Number of Open Auto Finance Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.10	90.0%	84.1%	88.0%	85.3%
0 or more	1.00	10.0%	15.9%	12.0%	14.7%

8) Average Number of Months Account have been Open

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 32	0.68	3.8%	6.4%	9.8%	9.1%
32 - 75	0.90	30.5%	42.5%	45.2%	40.5%
75 - 118	0.95	41.7%	34.8%	32.7%	37.2%
118 or more	1.00	24.0%	16.4%	12.3%	13.2%

9) Number of Open Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 12	1.10	81.3%	76.0%	76.8%	75.4%
12 or more	1.00	18.7%	24.0%	23.2%	24.6%

(continued. . .)

TABLE 11.
Credit-Based Insurance Scoring Model Developed by the FTC Using a Sample of Only Non-Hispanic White Insurance Customers (Continued)

10) *Variable H: Presence of a Particular Kind of Delinquency on the Account*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.28	98.8%	95.4%	97.7%	98.7%
1 or more	1.00	1.2%	4.6%	2.3%	1.3%

11) Ratio of Open Personal Financial Accounts to Total Open Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.90	82.0%	66.4%	73.6%	82.9%
0 or more	1.00	18.0%	33.6%	26.4%	17.1%

12) *Variable D: Presence of a Particular Kind of Delinquency on the Account*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.16	83.9%	54.1%	68.9%	82.3%
1 or more	1.00	16.1%	45.9%	31.1%	17.7%

13) *Variable C: Presence of Delinquencies on a Particular Kind of Account*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.90	25.1%	23.8%	20.0%	25.1%
0	0.81	71.8%	67.0%	72.5%	71.4%
1 or more	1.00	3.1%	9.2%	7.5%	3.5%

(continued. . .)

TABLE 11.
Credit-Based Insurance Scoring Model Developed by the FTC Using a Sample of Only Non-Hispanic White Insurance Customers (Continued)

14) *Variable I: Relates to the Number of Accounts in the Credit File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
Disputed	1.41	0.2%	0.5%	0.5%	0.0%
0 - 2	0.85	2.2%	5.0%	3.3%	2.4%
2 or more	1.00	97.6%	94.6%	96.2%	97.5%

15) Number of Bank Installment Accounts Ever Bad Debt

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.37	44.0%	46.8%	48.0%	49.9%
0	1.37	54.7%	49.5%	49.6%	48.7%
1 or more	1.00	1.3%	3.7%	2.4%	1.3%

Notes:
1) Variables in italics have not been described publicly, and ChoicePoint considers the descriptions of those variables to be proprietary information.

2) This scoring model was developed to use credit history information to predict the relative risk posed by individuals, where risk is defined as expected total dollars that would be paid out on claims in a year. To calculate a score for a given individual with this model, the appropriate factors for each of the 15 variables are multiplied together. The resulting product is the inverse of the estimated relative riskiness of the individual, based on the individual's credit history. This scoring model was developed using a development sample of only non-Hispanic white insurance customers. See Appendix E for a detailed discussion of the score-building process.

TABLE 12.
Credit-Based Insurance Scoring Model Developed by the FTC by Discounting Variables with Large Differences Across Racial and Ethnic Groups

1) *Variable J: Indebtedness on Accounts of a Particular Type*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.22	5.6%	10.0%	8.9%	4.9%
\$0 - \$1,000	1.34	36.0%	28.6%	34.4%	38.1%
\$1000 - \$3,000	1.25	20.1%	18.0%	17.9%	20.8%
\$3,000 - \$14,000	1.14	27.4%	31.9%	29.5%	25.5%
\$14,000 or more	1.00	10.9%	11.5%	9.2%	10.7%

2) *Variable E: Relates to the Number of Accounts in the Credit File*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.72	0.2%	0.4%	0.5%	0.0%
0 - 3	0.80	3.7%	5.6%	4.6%	3.7%
3 or more	1.00	96.2%	94.1%	94.9%	96.3%

3) Share of all Accounts that are Open

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - .14	0.83	2.1%	2.6%	2.3%	1.6%
.14 - .27	0.89	8.3%	9.1%	8.1%	8.7%
.27 or more	1.00	89.6%	88.4%	89.6%	89.7%

4) Number of Open Auto Finance Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.26	90.0%	84.1%	88.0%	85.3%
0 or more	1.00	10.0%	15.9%	12.0%	14.7%

(continued. . .)

TABLE 12.
Credit-Based Insurance Scoring Model Developed by the FTC by Discounting Variables with Large Differences Across Racial and Ethnic Groups (Continued)

5) Number of Open Bank Installment Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.13	68.5%	68.5%	67.9%	69.4%
0 or more	1.00	31.5%	31.5%	32.1%	30.6%

6) Number of Open Oil Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.85	91.6%	93.7%	88.7%	91.2%
0 or more	1.00	8.4%	6.3%	11.3%	8.8%

7) Ratio of Open Oil Accounts to Total Open Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.00	91.6%	93.7%	88.7%	91.2%
0 - .0741	0.86	4.6%	4.3%	6.2%	5.3%
.0741 or more	1.00	3.8%	2.1%	5.2%	3.5%

8) Number of Accounts Opened in the Last 3 Months

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0	1.15	79.3%	75.6%	74.7%	74.5%
1 or more	1.00	20.7%	24.4%	25.3%	25.5%

(continued. . .)

TABLE 12.
Credit-Based Insurance Scoring Model Developed by the FTC by Discounting Variables with Large Differences Across Racial and Ethnic Groups (Continued)

9) Number of Credit Union Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.06	36.1%	37.2%	34.0%	31.7%
1 - 5	1.06	51.5%	49.0%	52.3%	56.0%
5 or more	1.00	12.4%	13.7%	13.6%	12.3%

10) Age of Last Activity

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
0 - 2	0.88	98.3%	97.8%	98.5%	98.5%
2 or more	1.00	1.7%	2.2%	1.5%	1.5%

11) *Variable K: Number of Accounts of a Particular Type*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.77	5.4%	9.8%	8.9%	4.8%
0 - 6	0.96	77.8%	75.2%	74.6%	68.3%
6 or more	1.00	16.8%	15.0%	16.6%	26.9%

12) Ratio of Open Department Store Accounts to Total Open Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.99	32.0%	31.7%	28.1%	33.2%
0 - .36	0.93	58.9%	59.0%	61.6%	60.0%
.36 or more	1.00	9.1%	9.3%	10.3%	6.8%

(continued. . .)

TABLE 12.
Credit-Based Insurance Scoring Model Developed by the FTC by Discounting Variables with Large Differences Across Racial and Ethnic Groups (Continued)

13) Ratio of Open Bank Installment Accounts to Total Open Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	1.00	68.5%	68.5%	67.9%	69.4%
0 - .2917	1.10	27.5%	28.3%	28.6%	27.2%
.2917 or more	1.00	4.0%	3.2%	3.5%	3.4%

14) *Variable L: Based on Total Available Credit*

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
$0 - $3,000	0.91	4.7%	6.8%	6.7%	4.6%
$3,000 or more	1.00	95.3%	93.2%	93.3%	95.4%

15) Ratio of Open Credit Union Accounts to Total Open Accounts

		Share in each category, by race or ethnicity			
Category	Factor	Non-Hispanic Whites	African Americans	Hispanics	Asians
No trade line of this type	0.97	49.8%	48.9%	46.3%	45.1%
0 - .0789	0.95	7.4%	8.9%	8.6%	8.6%
.0789 or more	1.00	42.8%	42.2%	45.1%	46.3%

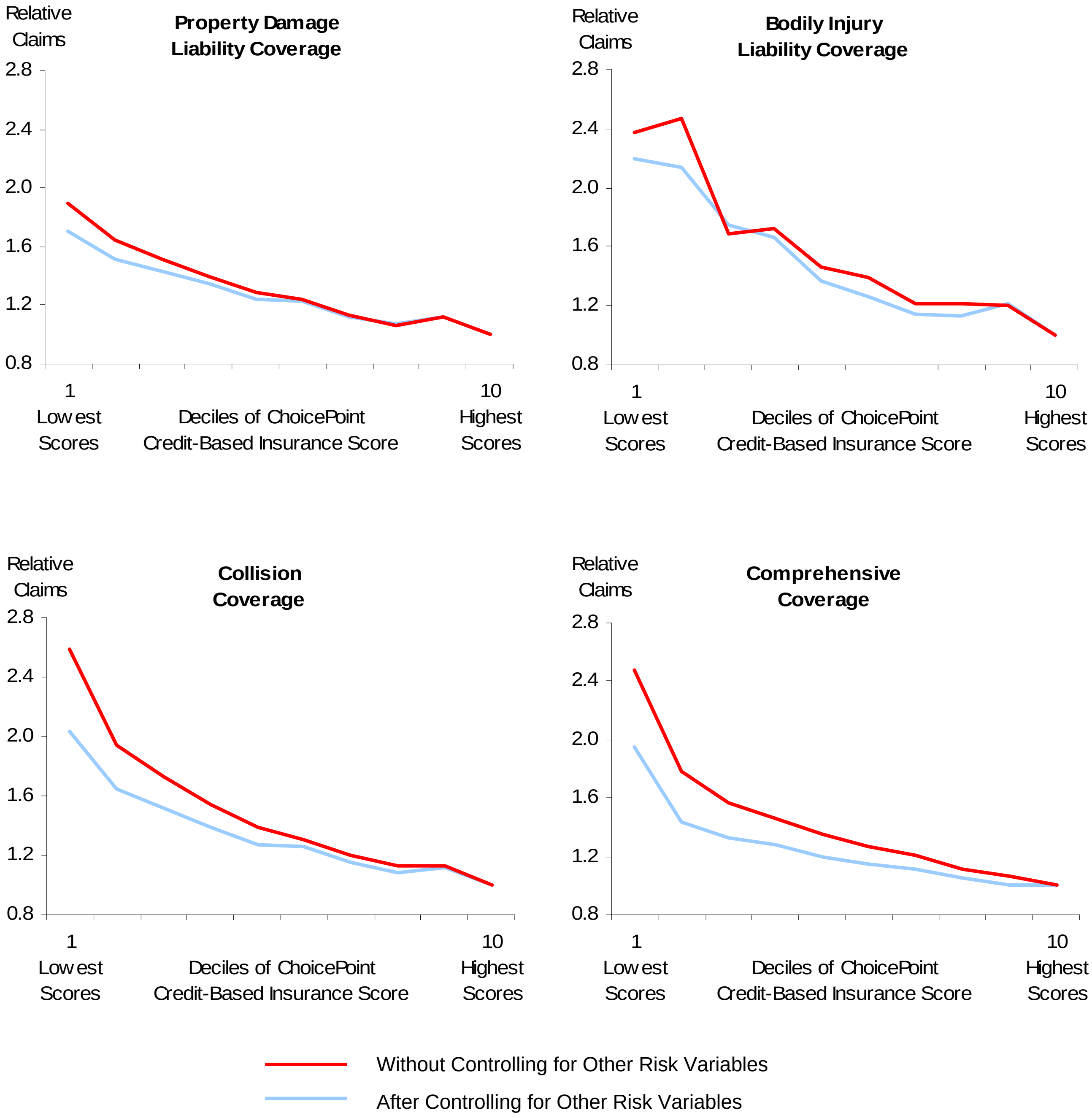
Notes:
1) Variables in italics have not been described publicly, and ChoicePoint considers the descriptions of those variables to be proprietary information.

2) This scoring model was developed to use credit history information to predict the relative risk posed by individuals, where risk is defined as expected total dollars that would be paid out on claims in a year. To calculate a score for a given individual with this model, the appropriate factors for each of the 15 variables are multiplied together. The resulting product is the inverse of the estimated relative riskiness of the individual, based on the individual's credit history. This scoring model was developed by discounting the predictive power of variables that had large differences across racial and ethnic groups, so that those variables would be less likely to be chosen by the score-building procedure. See Appendix E for a detailed discussion of the score-building process.

FIGURES

FIGURE 1.

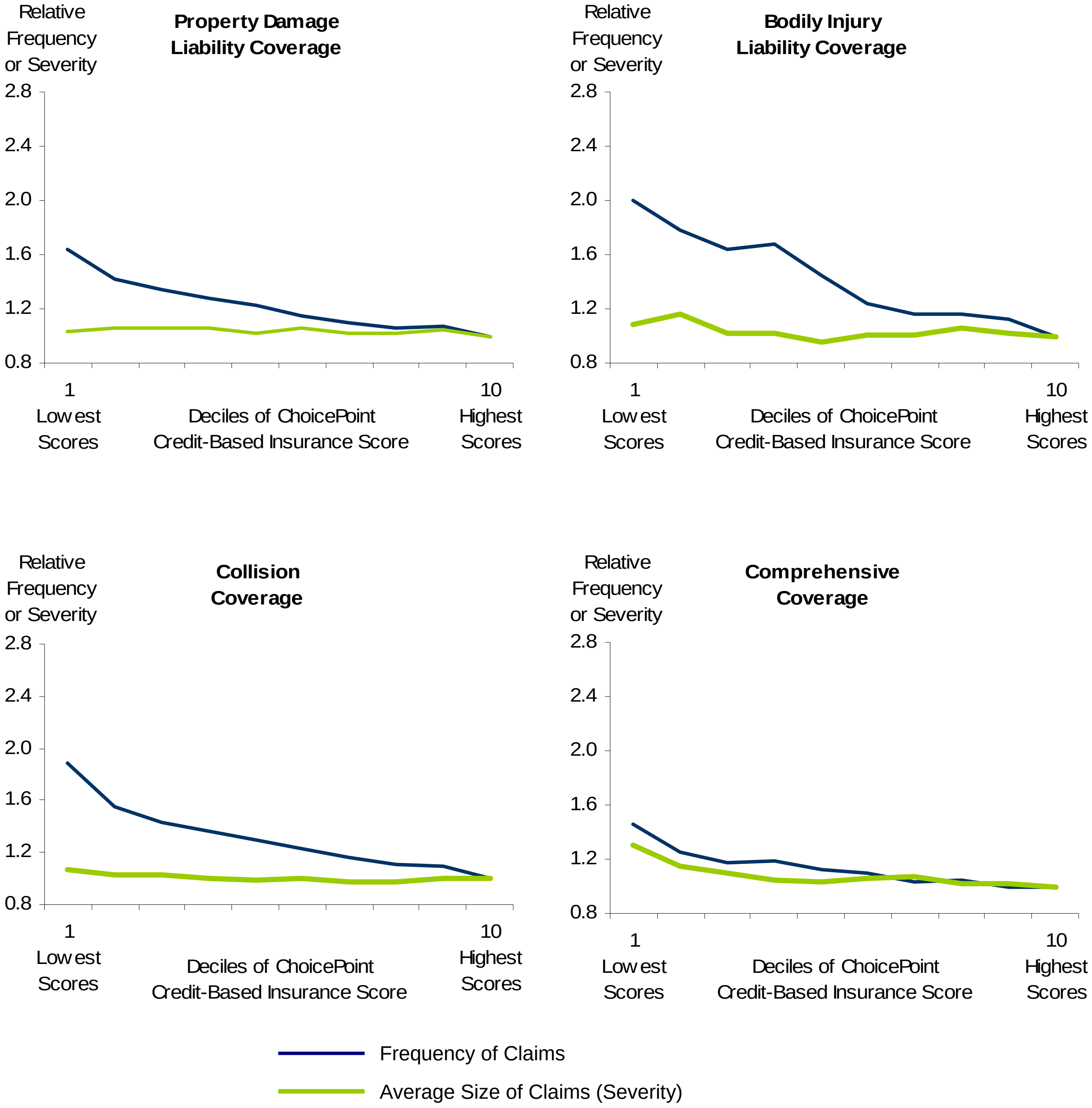
Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 2.

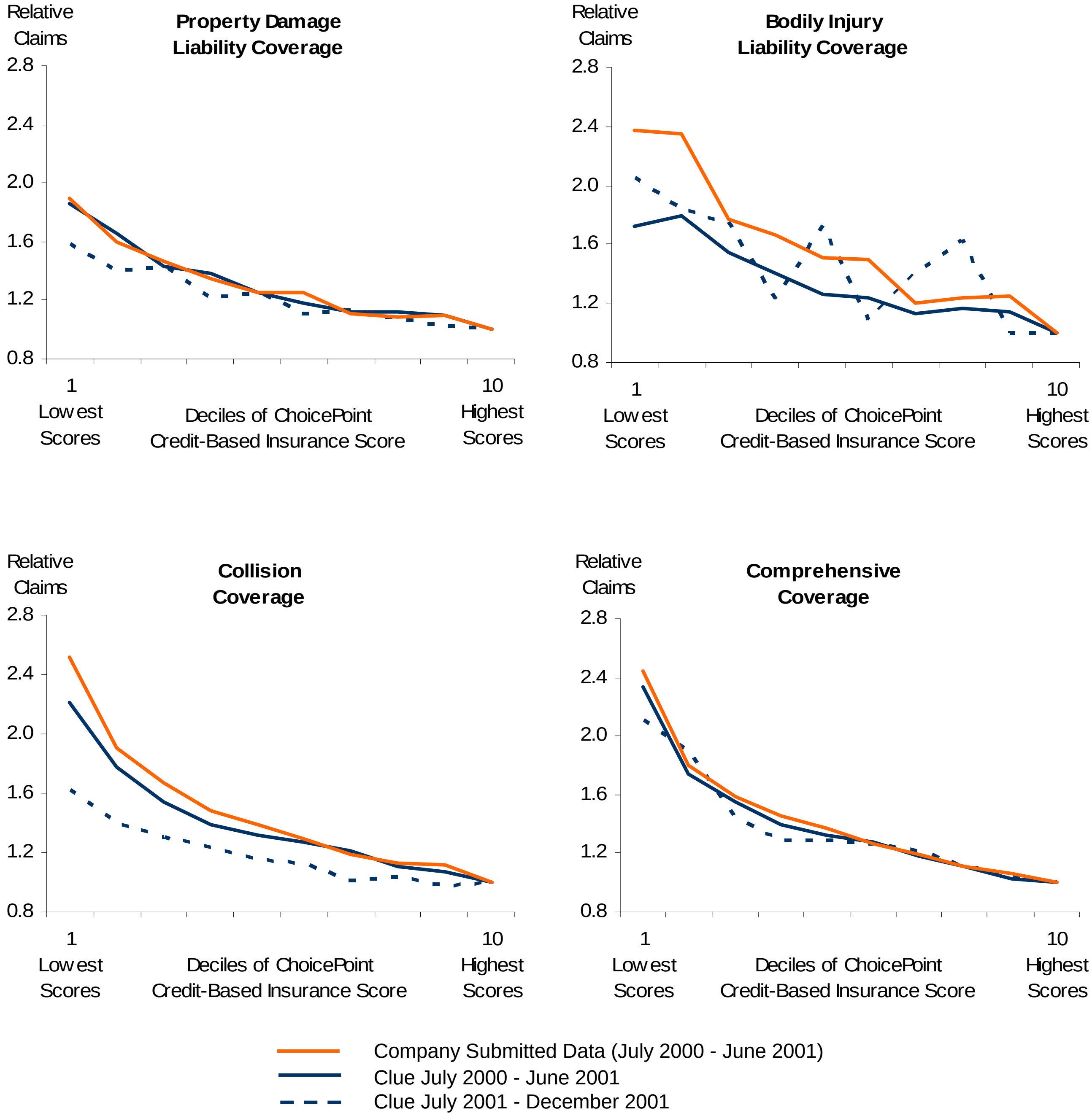
Frequency and Average Size (Severity) of Claims, Relative to Highest Score Decile



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 3.

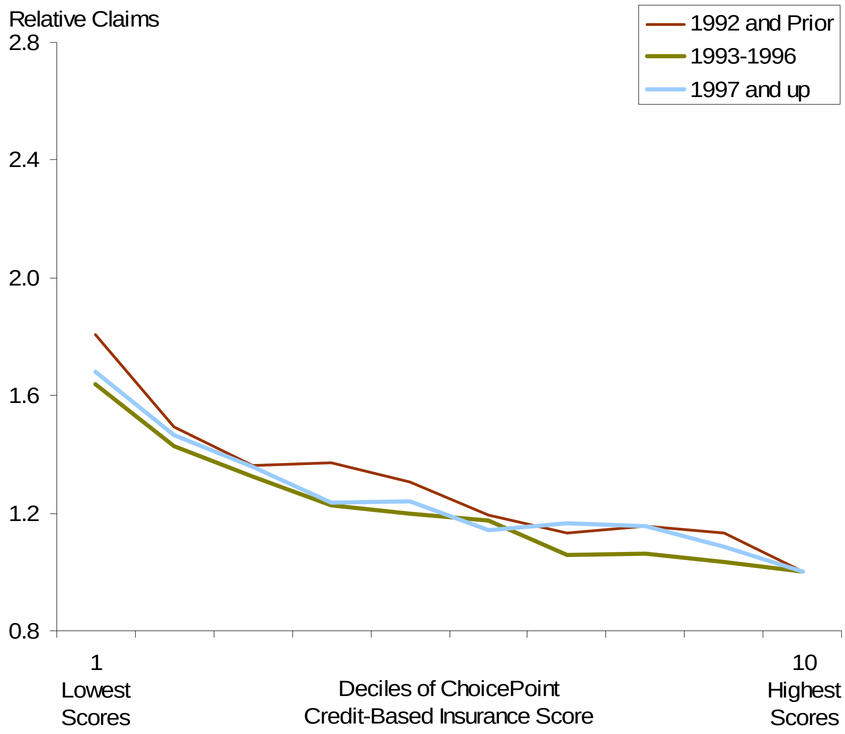
"CLUE" Claims Data: Average Amount Paid Out on Claims, Relative to Highest Score Decile



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 4.

By Model Year of Car: Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile (Property Damage Liability Coverage)

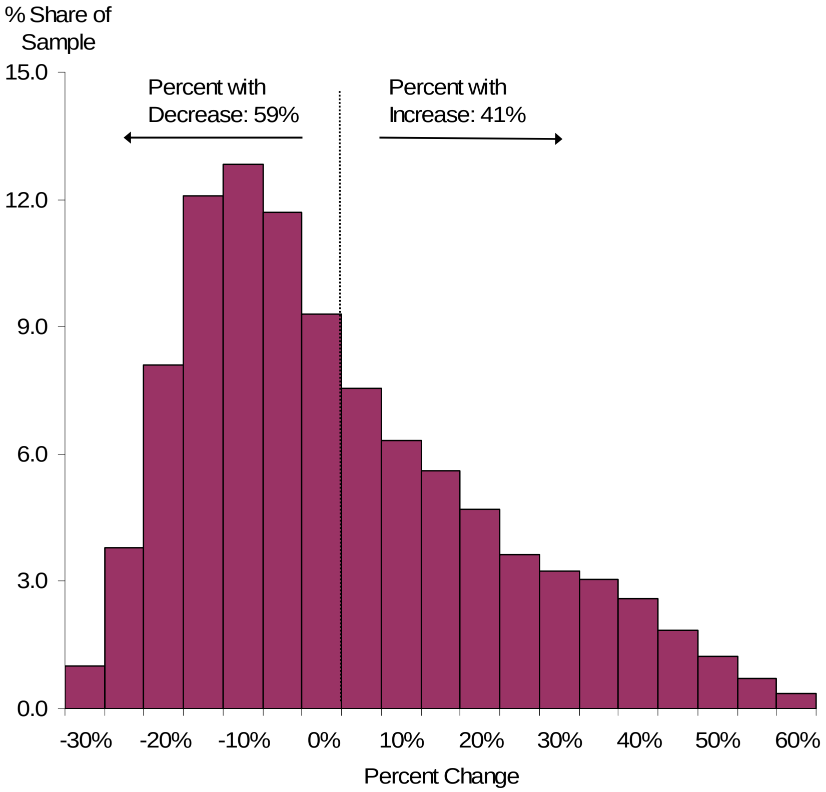


See notes on Figures at the end of this section.

Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 5.

Change in Predicted Amount Paid on Claims from Using Scores

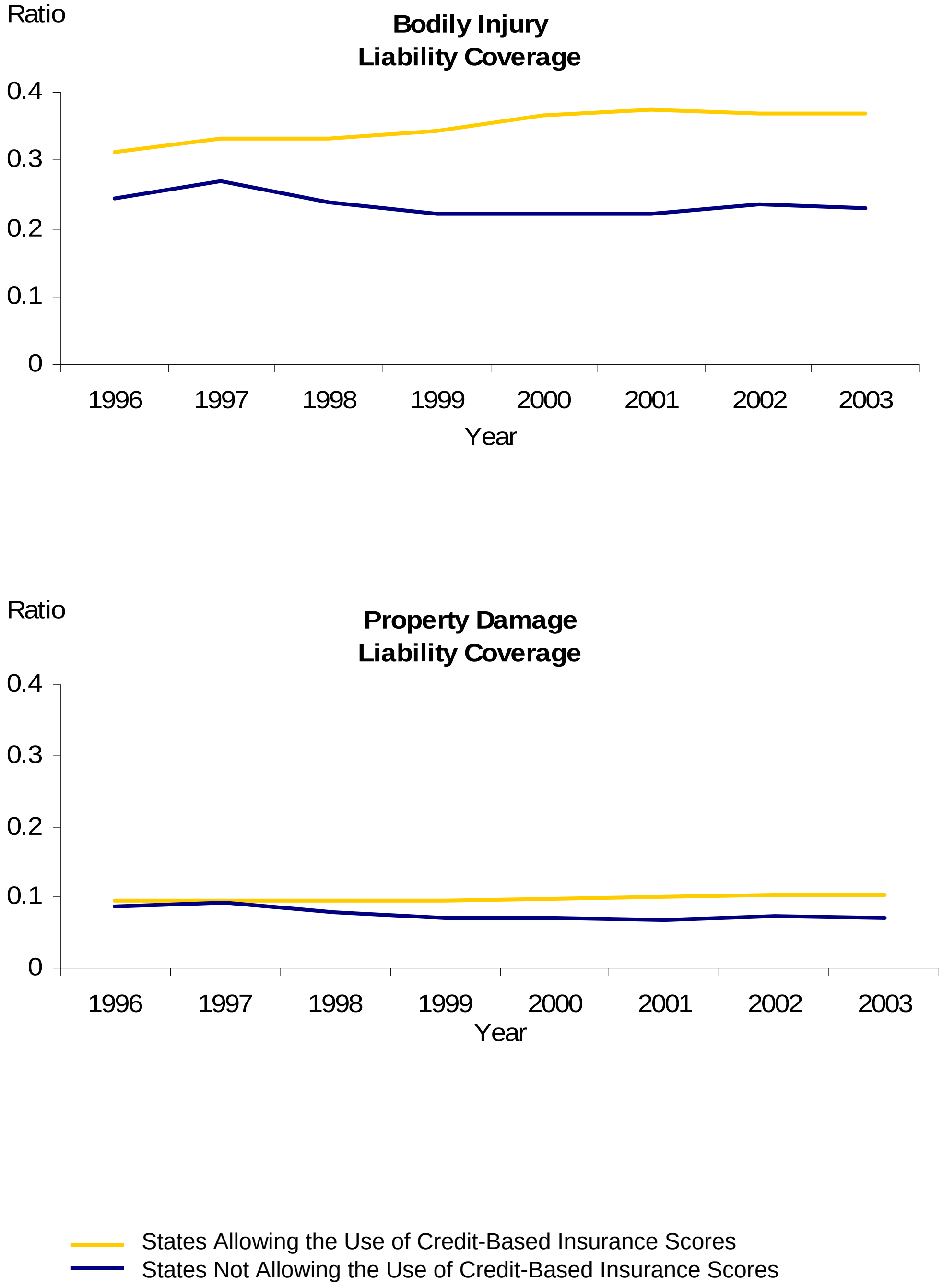


See notes on Figures at the end of this section.

Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 6.

The Ratio of Uninsured Motorist Claims to Liability Coverage Claims (1996-2003)

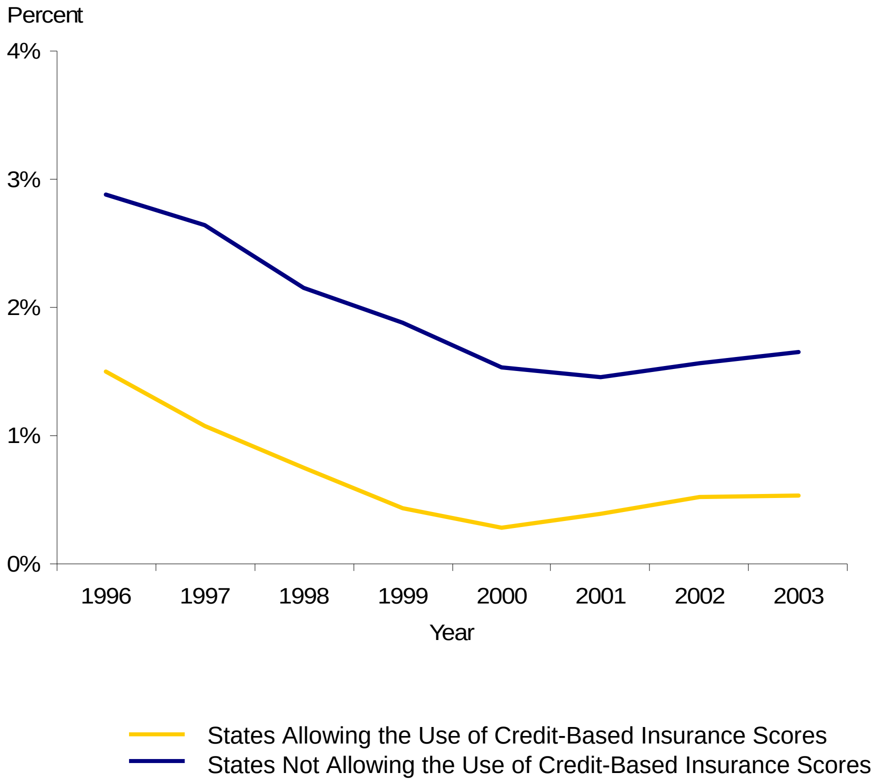


See notes on Figures at the end of this section.

Source: Analysis of data from several National Association of Insurance Commissioners Database Reports.

FIGURE 7.

Share of Cars Insured through States' "Residual Market" Insurance Programs (1996-2003)



See notes on Figures at the end of this section.

Source: Analysis of data from several National Association of Insurance Commissioners Database Reports.

FIGURE 8.

Distribution of Scores, by Race and Ethnicity

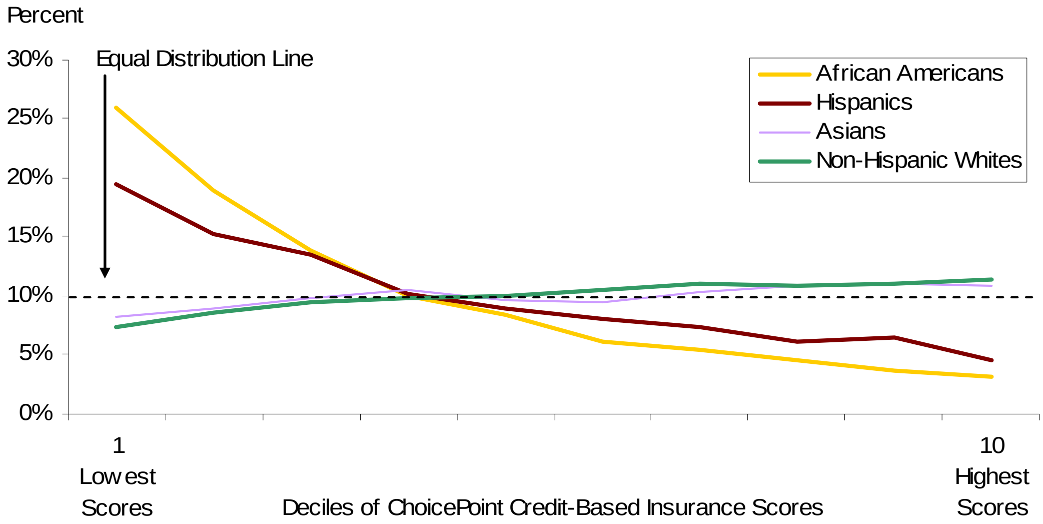
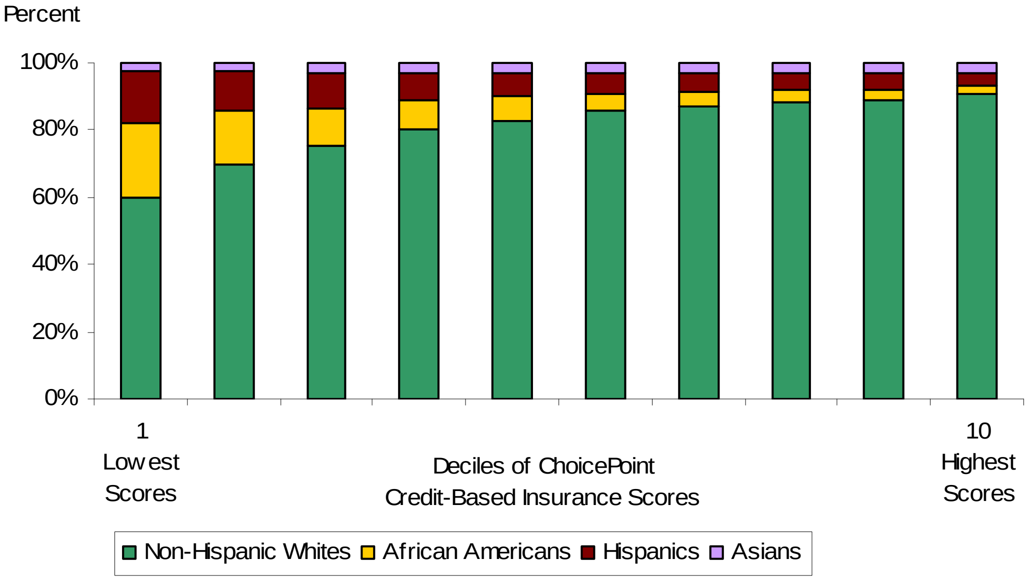


FIGURE 9.

Distribution of Race and Ethnicity, by Score Decile



See notes on Figures at the end of this section.

Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 10.

Distribution of Scores, by Neighborhood Income

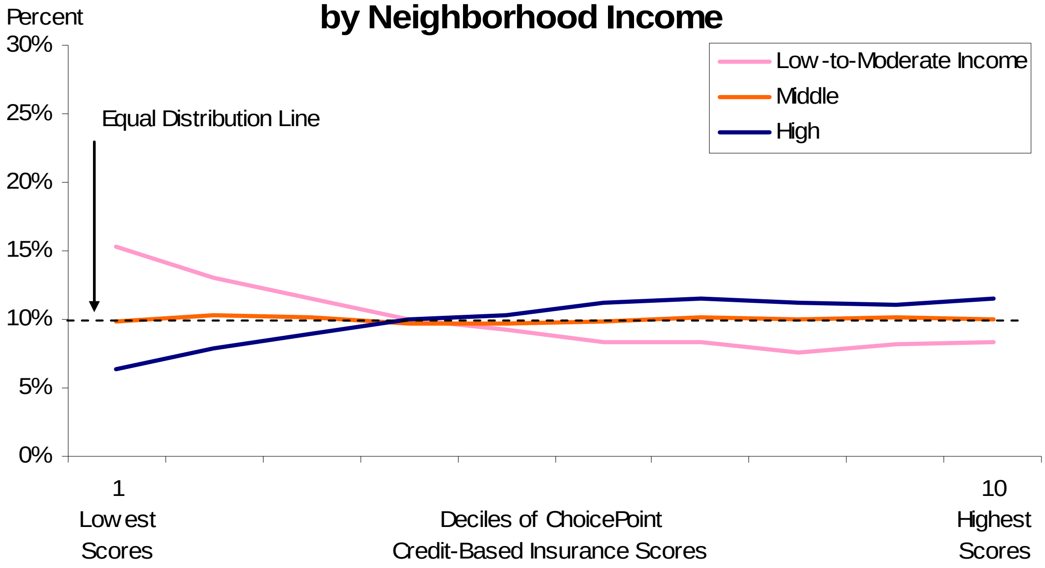
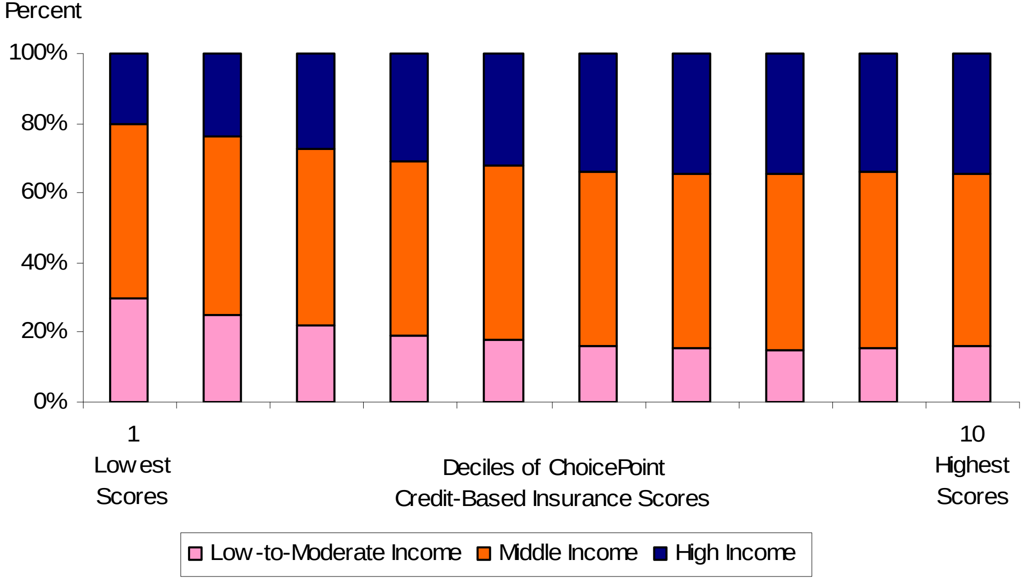


FIGURE 11.

Distribution of Neighborhood Income, by Score Decile

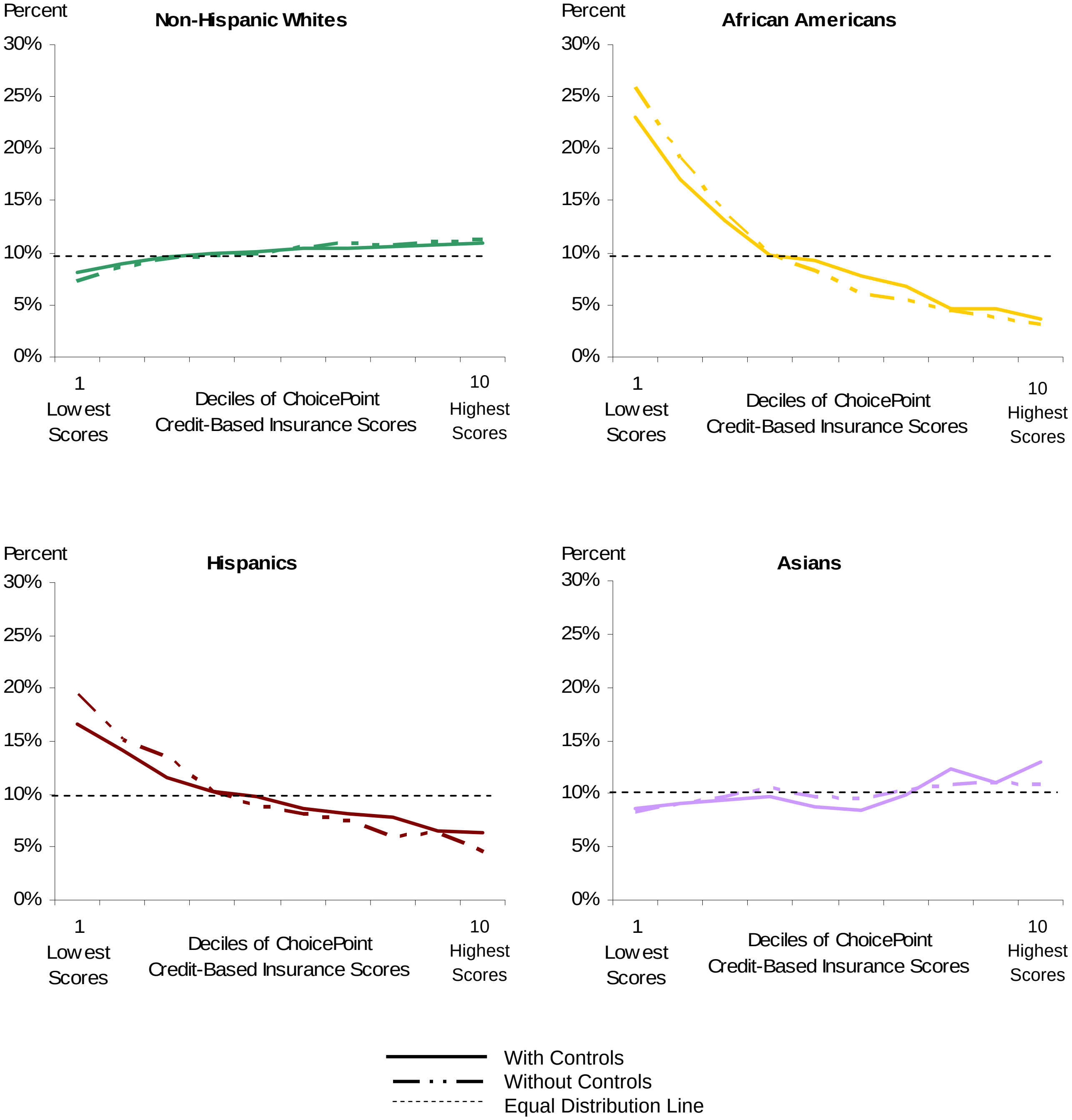


See notes on Figures at the end of this section.

Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 12.

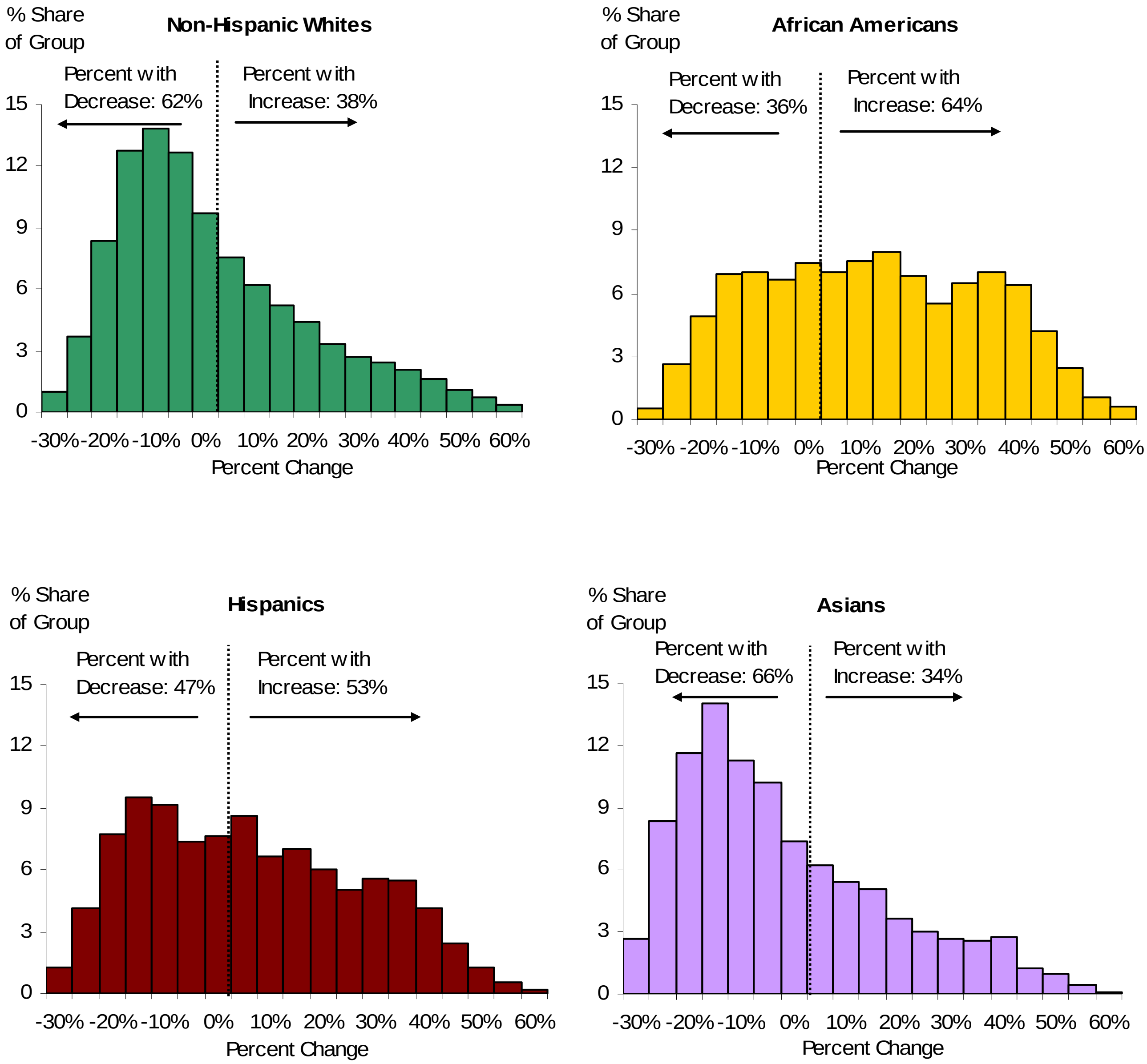
Distribution of Scores by Race and Ethnicity, After Controlling for Age, Gender, and Neighborhood Income



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 13.

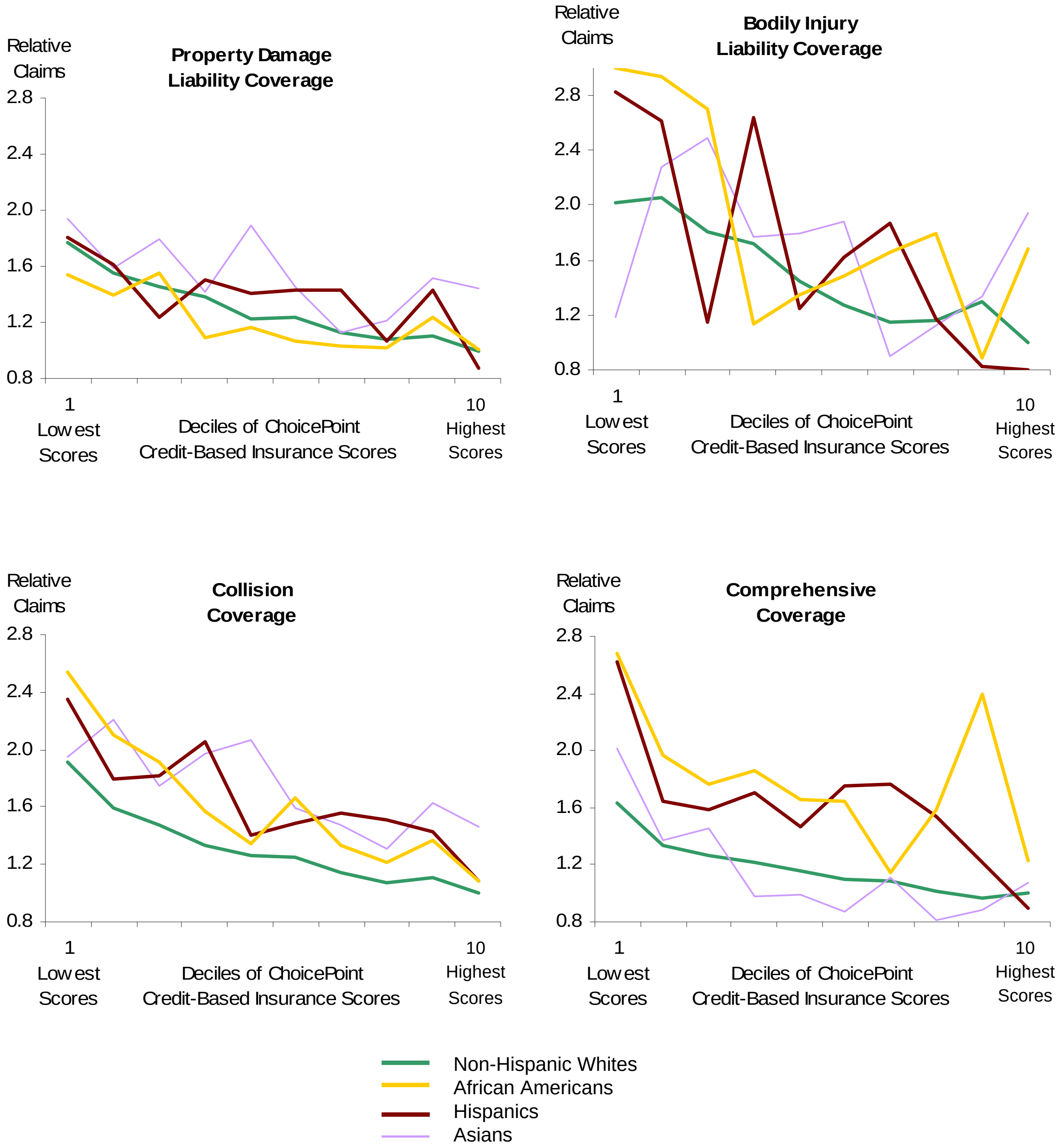
By Race and Ethnicity: Change in Predicted Amount Paid on Claims from Using Scores



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 14.

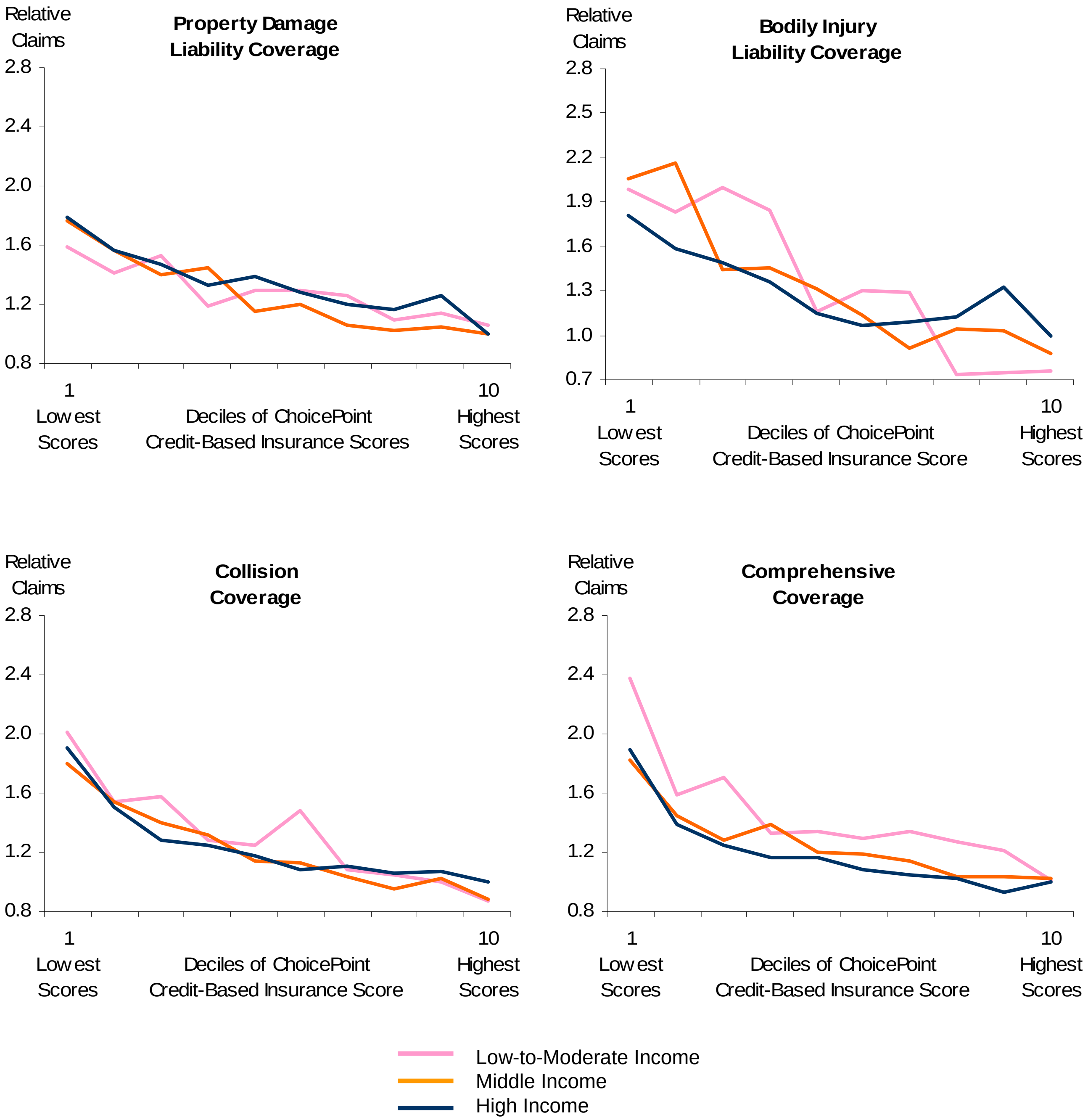
By Race and Ethnicity:
Estimated Average Amount Paid Out on Claims, Relative to Non-Hispanic Whites in Highest Score Decile



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 15.

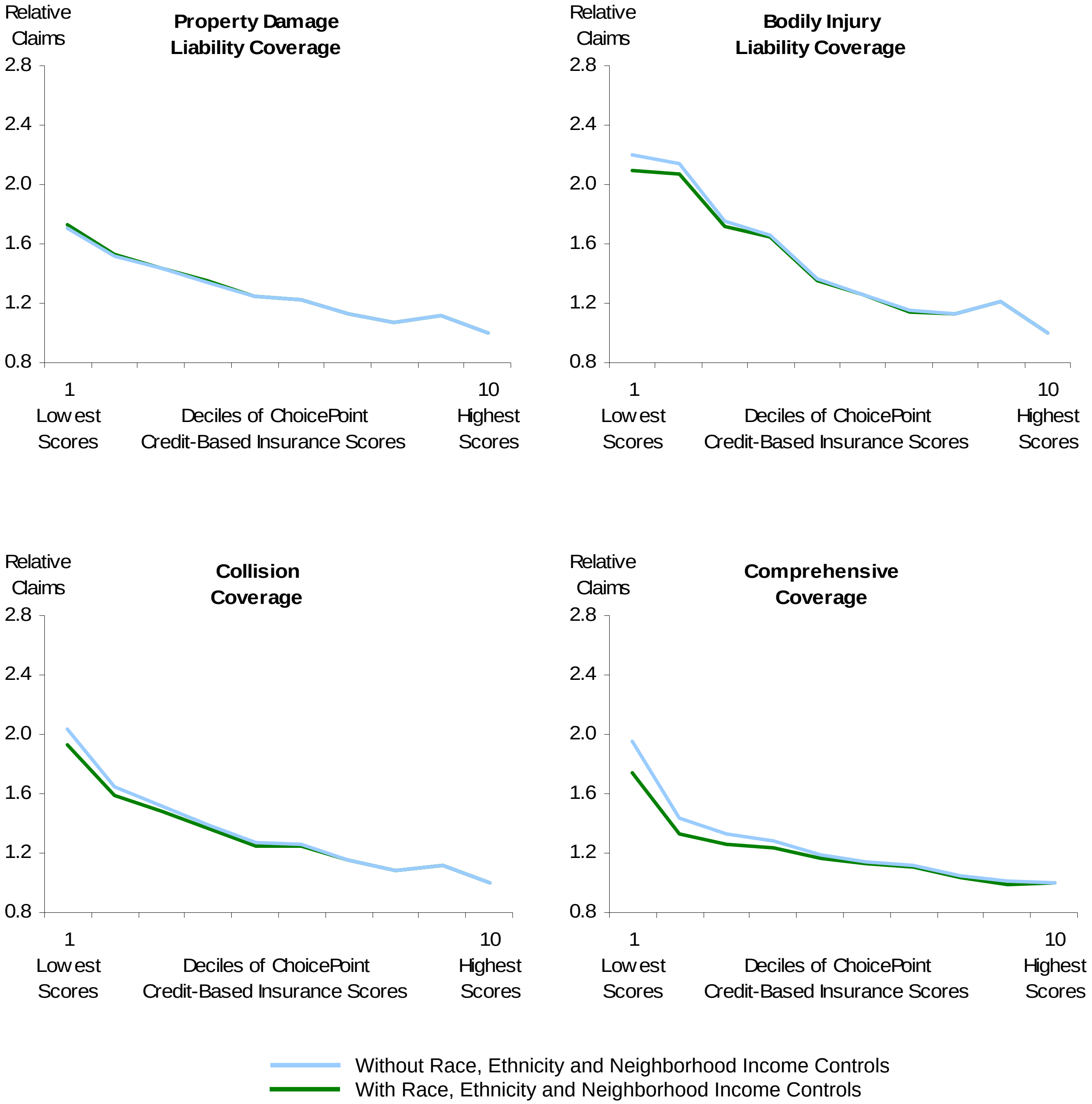
By Neighborhood Income: Estimated Average Amount Paid Out on Claims, Relative to People in Highest Score Decile in High Income Areas



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 16.

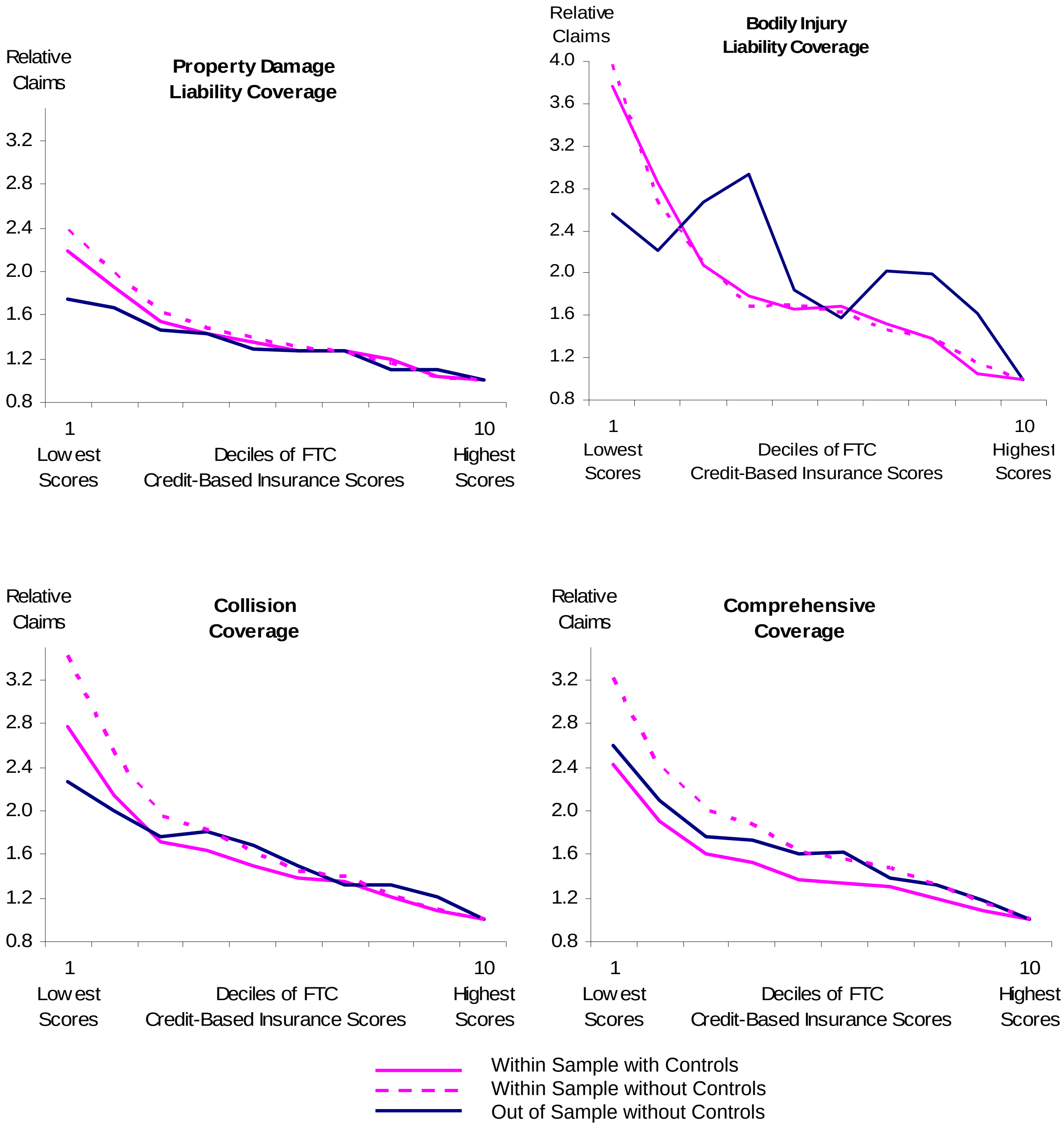
Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile, with and without Controls for Race, Ethnicity, and Neighborhood Income



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 17.

FTC Baseline Model - Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

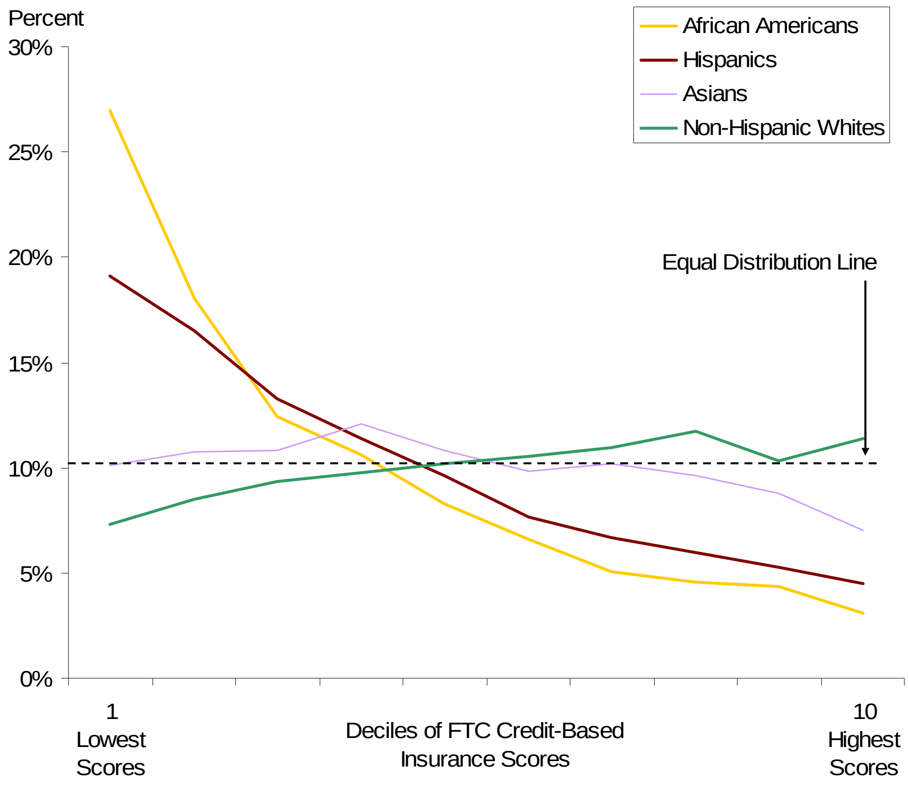


Note that the vertical scale on these graphs is different than for previous graphs of relative claims and score deciles.

See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 18.

Distribution of FTC Baseline Model Credit-Based Insurance Scores, by Race and Ethnicity

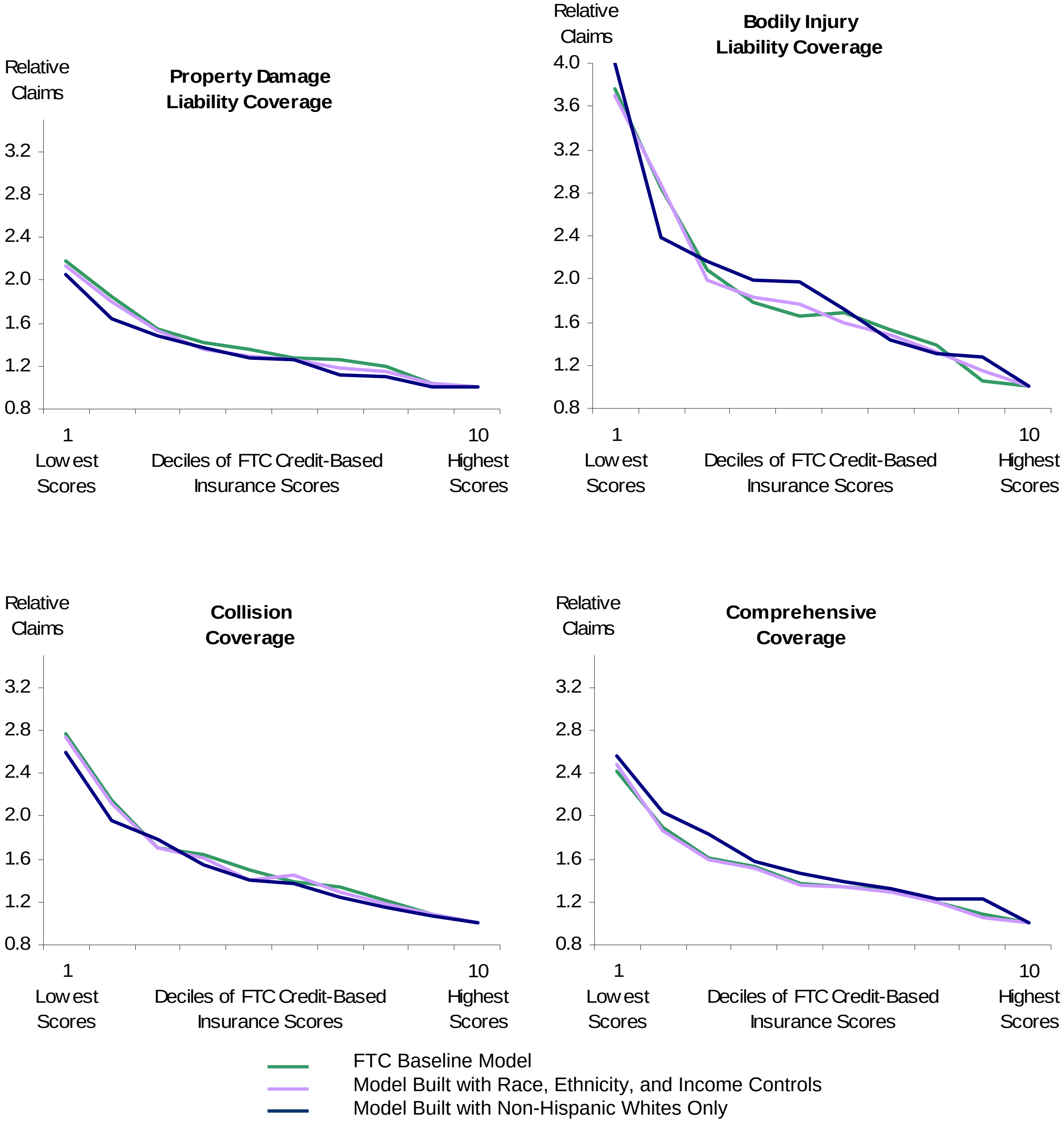


See notes on Figures at the end of this section.

Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 19.

FTC Score Models Built Controlling for Race, Ethnicity, and Neighborhood Income: Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

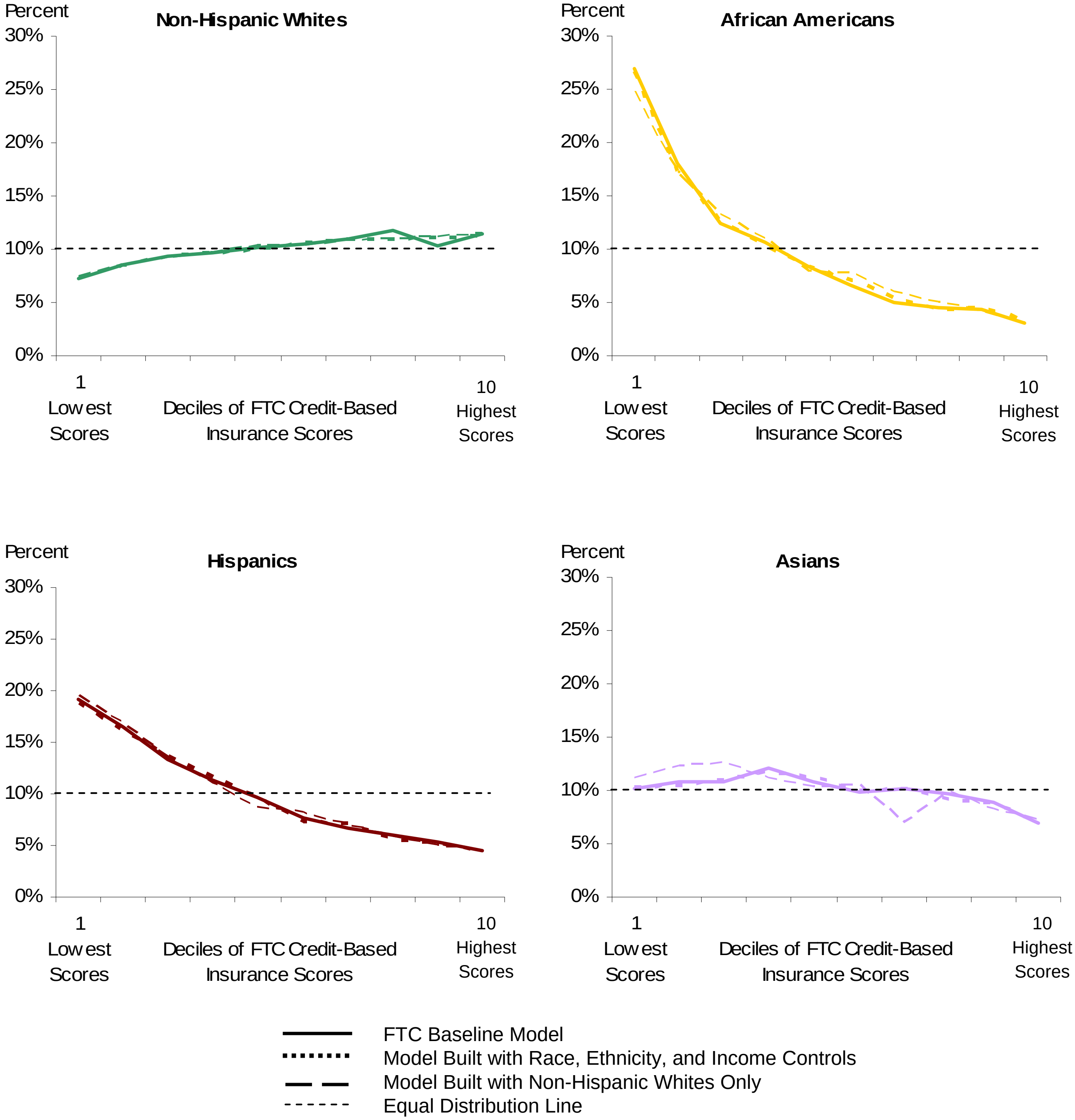


Note that the vertical scale on these graphs is different than for some previous graphs of relative claims and score deciles.

See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 20.

Distribution of FTC Credit-Based Insurance Scores, by Race and Ethnicity (A)

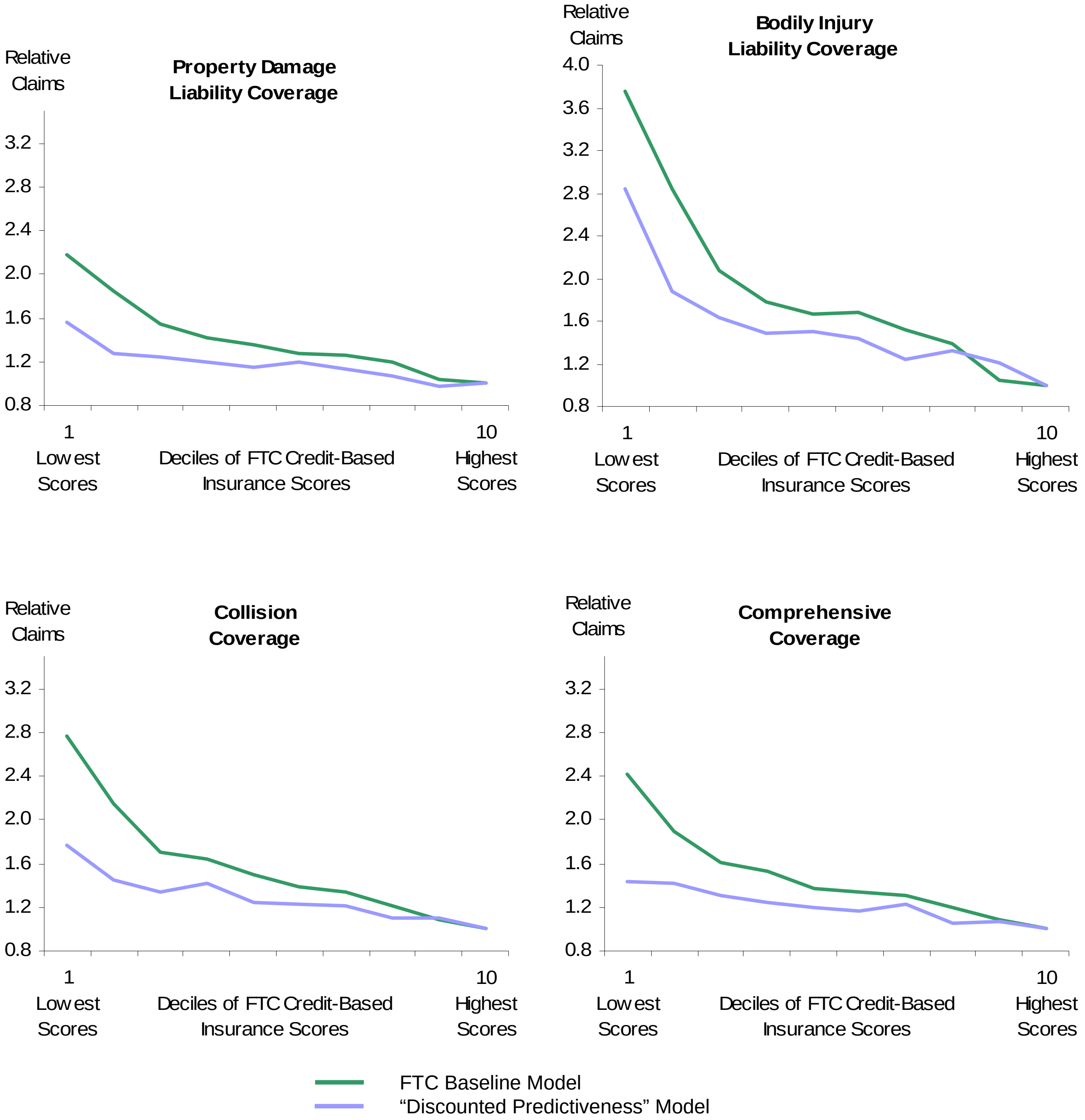


See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 21.

An Additional FTC Credit-Based Insurance Scoring Model: The "Discounted Predictiveness" Model

Estimated Average Amount Paid Out on Claims, Relative to Highest Score Decile

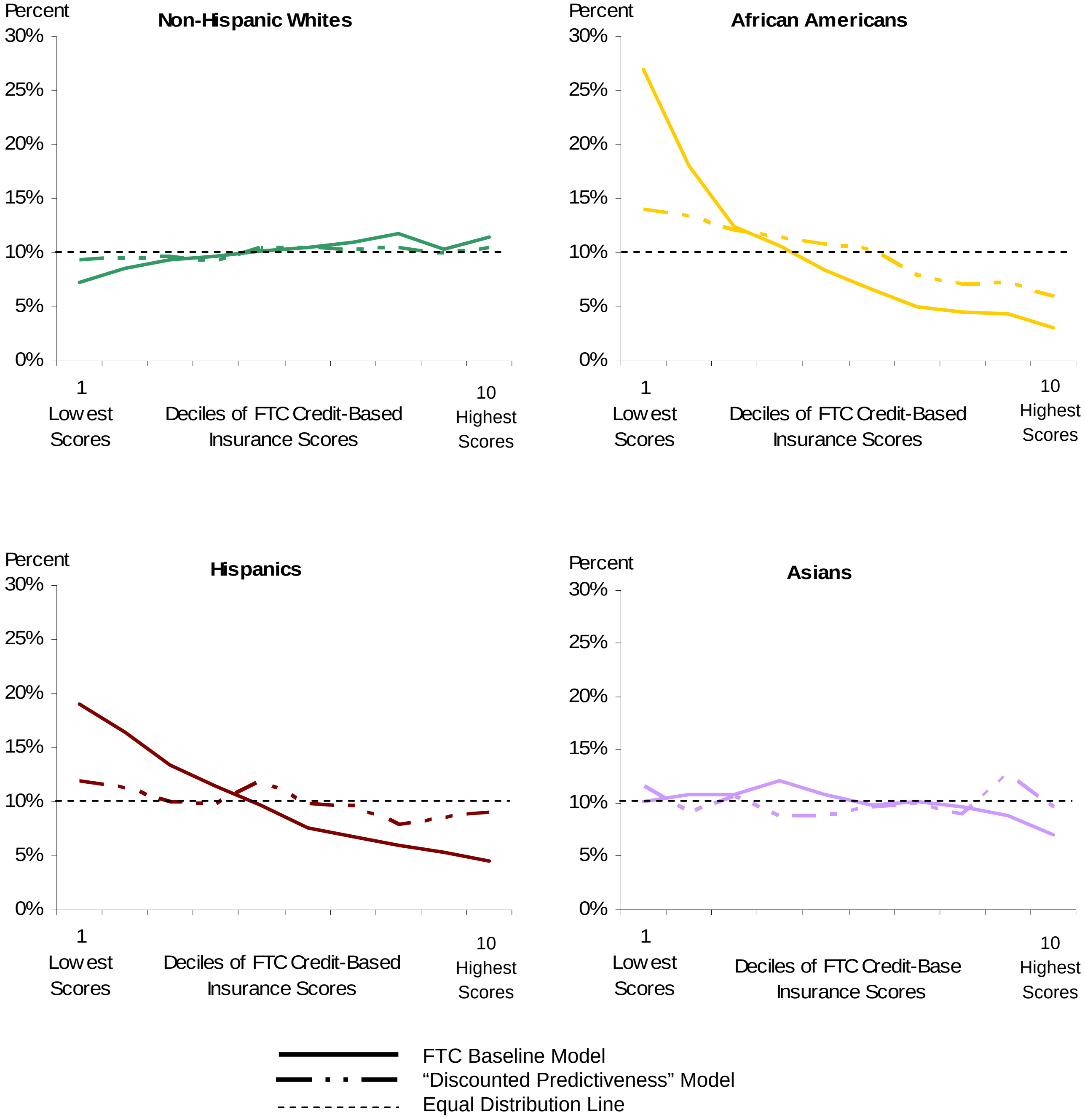


Note that the vertical scale on these graphs is different than for some previous graphs of relative claims and score deciles.

See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

FIGURE 22.

Distribution of FTC Credit-Based Insurance Scores, by Race and Ethnicity (B)



See notes on Figures at the end of this section.
Source: Analysis of FTC Automobile Insurance Policy Database

Notes on Figures

Figure 1:

The lines labeled "without controlling for other variables" show the actual average amount paid out on claims per year of coverage for each score decile, relative to the highest score decile. These are derived from the information in Table 2. For example, the relativity for the lowest decile on the PD graph has a value of 1.89. This number is calculated from column (c) on Table 2; by taking the average total paid on PD claims per year of coverage for the 1st decile ($118.73) and dividing it by the respective value for the 10th decile ($62.70).

The lines labeled "after controlling for other variables" show the predicted amount paid out on claims per year of coverage for each score decile, relative to the highest score decile, from Tweedie GLMs (Generalized Linear Models) of claims risk that included score and a set of standard risk variables as controls. Since our GLM models are multiplicative, the relativities shown by these lines are equivalent to the exponentiated coefficients of the score decile indicator variables. Modeling details and a description of the variables in the models are provided in Appendix D.

Figure 2:

The lines labeled "frequency of claims" show the predicted number of claims per year of coverage for each score decile, relative to the highest score decile, from Poisson GLM models ("Poisson Regressions") that included score and a set of standard risk variables as controls. Since our GLM models are multiplicative, the relativities shown by these lines are equivalent to the exponentiated coefficients of the score decile indicator variables. Modeling details and a description of the variables in the models are provided in Appendix D.

The lines labeled "average size of claims" show the predicted average size of claims for each score decile, relative to the highest score decile, from Gamma GLM models that included score and a set of standard risk variables as controls. Since our GLM model is multiplicative, the relativities shown by these lines are equivalent to the exponentiated coefficients of the score decile indicator variables. Modeling details and a description of the variables in the models are provided in Appendix D.

Figure 3:

"CLUE" stands for Comprehensive Loss Underwriting Exchange. This informational/ database exchange service is run by ChoicePoint, which collects data on claims from most major automobile insurance firms in the United States. These data allow firms to determine whether a potential new customer has filed a claim under a previous policy with another firm, and use that information in underwriting and rating.

Each line on this graph shows the average total amount paid out on claims per year of coverage for each score decile, relative to the highest decile. These results do not include controls for other risk variables because reliable non-credit risk variables are not available for the CLUE claims data. For this figure we use the full sample of 1.4 million policies, as opposed to the set of policies within the sub-sample of 400,000 normally used. This is because the latter would have proved a very limited sub-sample for the CLUE analysis for the half a year period moving forward, *i.e.*, for July 2001 to December 2001. See Appendix C for a description of the company-provided claims data and the CLUE database and claims data.

Figure 4:

Each line shows the predicted amount paid out on claims per year of coverage for each score decile, relative to the highest score decile, for each of three ranges of car model years from a Tweedie GLM risk model of claims that included score and a set of standard risk variables as controls. The different lines for the three groups of model years were estimated by interacting three model year range indicator variables with the score decile indicator variables. Modeling details and a description of the variables included in the models are provided in Appendix D.

Figure 5:

Predicted change in premium was estimated by comparing individuals' predicted total claims from risk models that included ChoicePoint Attract Standard Auto credit-based insurance score decile indicator variables with risk models that did not include scores. (By construction, the average of all changes is zero.) Both of these models were run separately for property damage liability, bodily injury liability, collision, and comprehensive coverage. In the final step we summed the predicted dollar risks for all four types of insurance coverage with and without the use of credit-based insurance scores. See section V.A. of the report for additional details on this analysis. Modeling details and a description of the variables included in the models are provided in Appendix D.

Figure 6:

Analysis based on data from several National Association of Insurance Commissioners Database Reports. (*e.g.*, National Association of Insurance Commissioners, “Auto Insurance Database Report 2003/2004” (2006)) The states included in the category “states not allowing the use of credit-based insurance scores” are California, New Jersey, Massachusetts, and Hawaii. The category "states allowing the use of credit-based insurance scores" includes all other states, except South Carolina and Texas (for which complete information was not provided in the NAIC reports).

Credit-based insurance scores for use in automobile insurance were first commercially available in 1995, and were widely adopted by insurance companies (in states that allowed their use) during the late 1990s.

Figure 7:

The “residual market” consists of state-sponsored programs to sell insurance to drivers who are unable to purchase insurance in the normal “voluntary” market. Analysis based on data from several National Association of Insurance Commissioners Database Reports. (*e.g.*, National Association of Insurance Commissioners, “Auto Insurance

Database Report 2003/2004" (2006)) The states included in the category "states not allowing the use of credit-based insurance scores" are California, New Jersey, Massachusetts, and Hawaii. The category "states allowing the use of credit-based insurance scores" includes all other states, except South Carolina and Texas (for which information was not provided in the NAIC report).

Credit-based insurance scores for use in automobile insurance were first commercially available in 1995, and were widely adopted by insurance companies (in states that allowed their use) during the late 1990s.

Figure 8:

Each line shows the share of each racial and ethnic group that is in each of the ten deciles of the ChoicePoint Attract Standard Auto credit-based insurance score. If each racial and ethnic group had the same distribution of scores, 10% of each group would be in each decile.

Figure 9:

[No Notes]

Figure 10:

Each line shows the share of each neighborhood income group that is in each of the ten deciles of the ChoicePoint Attract Standard Auto credit-based insurance score. If each neighborhood income group had the same distribution of scores, 10% of each group would be in each decile.

Figure 11:

[No Notes]

Figure 12:

Each line shows the share of each racial and ethnic group that is in each of the ten deciles of the ChoicePoint Attract Standard Auto credit-based insurance score *after controlling* for age, gender, and neighborhood income. This was calculated based on the residuals from an Ordinary Least Squares regression of ChoicePoint Attract Standard Auto credit-based insurance scores on age, gender, and neighborhood income. If each racial and ethnic group had the same distribution of scores, after controlling for age, gender, and neighborhood income, 10% of each group would be in each decile.

Figure 13:

Predicted change in premium was estimated by comparing individuals' predicted total claims from risk models that included ChoicePoint Attract Standard Auto credit-based insurance scores with risk models that did not include scores. By construction, the average of all changes for the entire sample is zero as in Figure 5, but the changes by race or ethnic group are not. See note for Figure 5 above or section V.A. of the report for additional details on this analysis. Modeling details and a description of the variables included in the models are provided in Appendix D.

Figure 14:

Each line shows the predicted amount paid out on claims per year of coverage for each score decile, relative to non-Hispanic whites in the highest score decile, from a Tweedie GLM risk model of claims that included score and a set of standard risk variables as controls. These values were generated by interacting the race and ethnicity indicator variables with the score decile indicator variables. The score decile cut-points used are the same across all race and ethnicity groups (these are the same deciles used for all previous Figures). Thus, given the race and ethnicity distributions across score deciles observed in Figure 8, there are relatively few African Americans and Hispanics in each of the higher score deciles intervals (*i.e.*, fewer than 10% of their group). Modeling details and a description of the variables included in the models are provided in Appendix D.

The differences in the estimates of the amount paid out in claims in higher score deciles versus the bottom score decile, within each race group, are generally statistically significant (at the 5% level), except for Asians (where they are only significant for comprehensive coverage). We also estimated the slope for each race and ethnicity group using a continuous score (as opposed to deciles), and found a statistically significant downward sloping relationship between score and the amount paid out in claims within each group, with the exception of bodily injury and property damage for Asians. Property damage for Asians did have a downward slope but was significant only at the 10% level. Note that Asians are the smallest race or ethnic group in our sample.

Figure 15:

Each line shows the predicted amount paid out on claims per year of coverage for each score decile, relative to the residents of high-income neighborhoods in the highest score decile, from a Tweedie GLM risk model of claims that included score and a set of standard risk variables as controls. These values were generated by interacting the neighborhood income category indicator variables with the score decile indicator variables. Modeling details and a description of the variables included in the models are provided in Appendix D.

Figure 16:

Each line shows the predicted amount paid out on claims per year of coverage for each score decile, relative to the highest score decile, from a Tweedie GLM risk model of claims that included score and a set of standard risk variables as controls. Since our GLM model is multiplicative, the relativities shown by this line are the exponentiated coefficients of the score decile indicator variables. The lines labeled “with race, ethnicity, and neighborhood controls” come from a model that also included indicator variables for race, ethnicity, and Census tract median income category. Modeling details and a description of the variables included in the models are provided in Appendix D.

Figure 17:

The line labeled “Within Sample” shows the predicted amount paid out on claims per year of coverage for each score decile relative to the highest score decile, *of the FTC baseline model*, from Tweedie GLM risk models of claims that included score and a set of standard risk variables as controls. Modeling details and a description of the variables included in the models are provided in Appendix D. Details on the score building process are provided in Appendix E.

The line labeled "Within Sample without Controls" shows the average total amount paid out on claims per year of coverage for each score decile relative to the highest decile, *of the FTC baseline model*, without controlling for any other risk variables. (This line is shown for comparison with the "Out of Sample" values below, for which we do not have controls.)

The "Out of Sample" line is based on CLUE claims data and shows the average total amount paid out on claims per year of coverage for each score decile relative to the highest decile, *of the FTC baseline model*, without controlling for any other risk variables (since reliable non-credit risk variables are not available in CLUE). This "Out of Sample" line is for the period July 2001 to December 2001, and uses CLUE claims data only for individuals who were not in the score development sample.

The development sample consisted only of the sub-sample of the FTC database for which we obtained SSA race and ethnicity data, which includes everyone who had a claim in the company data, so there is no way to use the company data to look at claims outside of the development sample. Therefore, we use CLUE data on claims for a different time period and for a different set of people instead (we were able to use data on roughly 800,000 policies for this from the original 1.4 million dataset). See Appendix C for a description of the CLUE database and claims data. Details on the score building process are provided in Appendix E.

(Note that the vertical scale on the graphs in this Figure rises higher than it does for previous graphs of relative claims and score deciles in Figures 1-4 and Figures 14-16)

Figure 18:

Each line shows the share of each racial and ethnic group that is in each of the ten deciles of the scores produced by the FTC's baseline credit-based insurance scoring model. If each racial and ethnic group had the same distribution of scores, 10% of each group would be in each decile.

Figure 19:

Each line shows the predicted amount paid out on claims per year of coverage for each score decile, relative to the highest score decile, from Tweedie GLM risk models of claims that included score and a set of standard risk variables as controls. Since our GLM models are multiplicative, the relativities shown by these lines are equivalent to the exponentiated coefficients of the score decile indicator variables. The lines labeled "baseline model" use scores from the FTC baseline scoring model. The lines labeled "race, ethnicity, and income controls model" use scores from a model built by controlling for those variables during the score building process. The lines labeled "Non-Hispanic whites model" come from a scoring model built using a development sample made up exclusively of non-Hispanic white insurance customers. Modeling details and a description of the variables included in the models are provided in Appendix D. Details on the score building process are provided in Appendix E.

(Note that the vertical scale on the graphs in this Figure rises higher than it does for previous graphs of relative claims and score deciles in Figures 1-4 and Figures 14-16)

Figure 20:

Each line shows the share of each racial and ethnic group that is in each of the ten deciles of three FTC credit-based insurance scoring models. The lines labeled "baseline model" use scores from the FTC baseline scoring model. The lines labeled "race, ethnicity, and income controls model" use scores from a model built by controlling for those variables during the score building process. The lines labeled "Non-Hispanic whites model" come from a scoring model built using a development sample made up exclusively of non-Hispanic white insurance customers. If each racial and ethnic group had the same distribution of scores, 10% of each group would be in each decile. Details on the score building process are provided in Appendix E.

Figure 21:

Each line shows the predicted relative amount paid out on claims per year of coverage for each score decile, relative to the highest score decile, from Tweedie GLM risk models of claims that included score and a set of standard risk variables as controls. Since our GLM models are multiplicative, the relativities shown by these lines are equivalent to the exponentiated coefficients of the score decile indicator variables. The lines labeled "baseline model" use scores from the FTC baseline scoring model. The lines labeled "discounted predictiveness model" use scores from a model built by discounting the power of a variable to predict risk based on how different the variable was across racial and ethnic groups. Modeling details and a description of the variables included in the models are provided in Appendix D. Details on the score building process are provided in Appendix E.

(Note that the vertical scale on the graphs in this Figure rises higher than it does for previous graphs of relative claims and score deciles in Figures 1-4 and Figures 14-16)

Figure 22:

Each line shows the share of each racial and ethnic group that is in each of the ten deciles of two FTC credit-based insurance scoring models. The lines labeled "baseline model" use scores from the FTC baseline scoring model. The lines labeled "discounted predictiveness model" use scores from a model built by discounting the power of a variable to predict risk based on how different the variable was across racial and ethnic groups. If each racial and ethnic group had the same distribution of scores, 10% of each group would be in each decile. Details on the score building process are provided in Appendix E.

APPENDIX A

TEXT OF SECTION 215 OF THE FACT ACT

SEC. 215. STUDY OF EFFECTS OF CREDIT SCORES AND CREDIT-BASEDINSURANCE SCORES ON AVAILABILITY AND AFFORDABILITYOF FINANCIAL PRODUCTS.

(a) STUDY REQUIRED.—The Commission and the Board, in consultation with the Office of Fair Housing and Equal Opportunity of the Department of Housing and Urban Development, shall conduct a study of—

(1) the effects of the use of credit scores and credit-based insurance scores on the availability and affordability of financial products and services, including credit cards, mortgages, auto loans, and property and casualty insurance;

(2) the statistical relationship, utilizing a multivariate analysis that controls for prohibited factors under the Equal Credit Opportunity Act and other known risk factors, between credit scores and credit-based insurance scores and the quantifiable risks and actual losses experienced by businesses;

(3) the extent to which, if any, the use of credit scoring models, credit scores, and credit-based insurance scores impact on the availability and affordability of credit and insurance to the extent information is currently available or is available through proxies, by geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, and creed, including the extent to which the consideration or lack of consideration of certain factors by credit scoring systems could result in negative or differential treatment of protected classes under the Equal Credit Opportunity Act, and the extent to which, if any, the use of underwriting systems relying on these models could achieve comparable results through the use of factors with less negative impact; and

(4) the extent to which credit scoring systems are used by businesses, the factors considered by such systems, and the effects of variables which are not considered by such systems.

(b) PUBLIC PARTICIPATION.—The Commission shall seek public input about the prescribed methodology and research design of the study described in subsection (a), including from relevant Federal regulators, State insurance regulators, community, civil rights, consumer, and housing groups.

(c) REPORT REQUIRED.—

(1) IN GENERAL.—Before the end of the 24-month period beginning on the date of enactment of this Act, the Commission shall submit a detailed report on the study conducted pursuant to subsection (a) to the Committee on Financial Services of the House of Representatives and the Committee on Banking, Housing, and Urban Affairs of the Senate.

(2) CONTENTS OF REPORT.—The report submitted under paragraph (1) shall

include the findings and conclusions of the Commission, recommendations to address specific areas of concerns addressed in the study, and recommendations for legislative or administrative action that the Commission may determine to be necessary to ensure that credit and credit-based insurance scores are used appropriately and fairly to avoid negative effects.

APPENDIX B

REQUESTS FOR PUBLIC COMMENT

[Billing Code 6750-01-P]

FEDERAL TRADE COMMISSION
RIN 3084- [AA94]

Public Comment on Methodology and Research Design for Conducting a Study of the Effects of Credit Scores and Credit-Based Insurance Scores on Availability and Affordability of Financial Products

AGENCY: Federal Trade Commission
ACTION: Notice and request for public comment.

SUMMARY: The Fair and Accurate Credit Transactions Act of 2003 ("FACT Act" or "Act") requires the Federal Trade Commission ("FTC" or "Commission") and the Federal Reserve Board ("Board") to conduct a study on the effects of credit scores and credit-based insurance scores on the availability and affordability of financial products. These products include credit cards, mortgages, auto loans, and property and casualty insurance. The Act requires the FTC to seek public input about "the prescribed methodology and research design of the study." As part of its efforts to fulfill its obligations under the Act, the FTC seeks public comment on how the FTC and the Board should conduct the study.

DATES: Comments must be received by August 16, 2004.

ADDRESSES: Public comments are invited, and may be filed with the Commission in either paper or electronic form. Comments should refer to "FACT Act Scores Study, Matter No. P044804," to facilitate their organization. A comment filed in paper form should include this reference both in the text and on the envelope, and should be mailed or delivered to: Federal Trade Commission/Office of the Secretary, Room H-159 (Annex N), 600 Pennsylvania Avenue, N.W., Washington, D.C. 20580. The FTC urges that any comment filed in paper form be sent by courier or overnight service, if possible, because

U.S. postal mail in the Washington area and at the Commission is subject to delay due to heightened security precautions.

Comments that do not contain any nonpublic information may be filed in electronic form (in ASCII format, WordPerfect, or Microsoft Word) as a part of or as an attachment to email messages directed to: FACTAscoringstudy@ftc.gov. If a comment contains nonpublic information, it must be filed in paper (rather than electronic) form, and the first page of the document must be clearly labeled "Confidential."[143]

The FTC Act and other laws the Commission administers permit the collection of public comments to consider and use in this proceeding as appropriate. All timely and responsive public comments, whether filed in paper or electronic form, will be considered by the Commission, and will be available to the public on the FTC Web site, to the extent practicable, at www.ftc.gov. As a matter of discretion, the FTC makes every effort to remove home contact information for individuals from the public comments it receives before placing those comments on the FTC Web site. More information, including routine uses permitted by the Privacy Act, may be found in the FTC's privacy policy, at http://www.ftc.gov/ftc/privacy.htm.

FOR FURTHER INFORMATION CONTACT: Jesse Leary, Deputy Assistant Director, (202) 326-3480, Division of Consumer Protection, Bureau of Economics, Federal Trade Commission, 600 Pennsylvania Avenue, N.W., Washington, DC 20580.

---

[143] 1 Commission Rule 4.2(d), 16 CFR 4.2(d). The comment must also be accompanied by an explicit request for confidential treatment, including the factual and legal basis for the request, and must identify the specific portions of the comment to be withheld from the public record. The request will be granted or denied by the Commission's General Counsel, consistent with applicable law and the public interest. See Commission Rule 4.9(c), 16 CFR 4.9(c).

SUPPLEMENTARY INFORMATION:

I. Background

The FACT Act was signed into law on December 4, 2003. Fair and Accurate Credit Transactions Act of 2003, Pub. L. No. 108-159 (2003). In general, the Act amends the Fair Credit Reporting Act ("FCRA") to enhance the accuracy of consumer reports and to allow consumers to exercise greater control regarding the type and amount of marketing solicitations they receive. To promote increasingly efficient national credit markets, the FACT Act also establishes uniform national standards in key areas of regulation regarding consumer report information. The Act contains a number of provisions intended to combat consumer fraud and related crimes, including identity theft, and to assist its victims. Finally, the Act requires a number of studies be conducted on credit reporting and related issues.

Section 215 of the FACT Act requires the FTC and the Board, in consultation with the Office of Fair Housing and Equal Opportunity of the Department of Housing and Urban Development, to conduct a study on the effects of credit scores and credit-based insurance scores on the availability and affordability of financial products. These products include mortgages, auto loans, credit cards, and property and casualty insurance. Section 215 further requires the FTC and the Board to study: 1) "the statistical relationship, utilizing a multivariate analysis that controls for prohibited factors under the Equal Credit Opportunity Act and other known risk factors, between credit scores and credit-based insurance scores and the quantifiable risks and actual losses;" and 2) "the extent to which, if any, the use of credit scoring models, credit scores, and credit-based insurance scores impact on the availability and affordability of credit to the extent information is currently available or is available through proxies, by geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, and creed, including the extent to which the consideration or lack of consideration of certain factors by credit scoring systems could result in negative or differential treatment of the protected classes, under the Equal Credit Opportunity Act, and the extent to which, if any, the use of underwriting systems relying on these models could achieve comparable results through the use of factors with less negative impact."

The study is due December 4, 2005.

II. Request for Comments

The Act requires the FTC to seek public input about "the prescribed methodology and research design of the study." As part of its efforts to fulfill its obligations under the Act, the FTC seeks public comment on how the FTC and the Board should conduct the study. Public comment is requested on all aspects of the study. In addition, the FTC seeks comment on the following questions:

1. How should the effects of credit scores and credit based insurance scores on the price and availability of mortgages, auto loans, credit cards, other credit products, and property and casualty insurance be studied? What is a reasonable methodology for measuring the price and availability of mortgages, auto loans, credit cards, other credit. Products, and property and casualty insurance, and the impact of credit scores and credit based insurance scores on those prices and availability?

2. An effect can often only be measured relative to a counterfactual (that is, relative to some hypothetical alternative situation). To determine the effects of credit scores on the price and availability of credit products, what is a reasonable counterfactual

to the current use of credit scores? To determine the effects of credit-based insurance scores on the price and availability of property and casualty insurance, what is a reasonable counterfactual to the current use of credit-based insurance scores?

3. Paragraph (a)(2) of Section 215 requires a study of "the statistical relationship, utilizing a multivariate analysis that controls for prohibited factors under the (ECOA) and other known risk factors, between credit scores and credit-based insurance scores and the quantifiable risks and actual losses experienced by businesses." (The ECOA "prohibited factors" are race, color, religion, national origin, sex or marital status, and age.) What is an appropriate multivariate technique for studying this relationship? What data would be required to undertake such an analysis? What data are available to undertake such an analysis?

4. What is an appropriate methodology to determine whether the use of credit scores or credit based insurance scores results in "negative or differential treatment" of ECOA-protected classes?

5. What is an appropriate methodology to determine whether the use of specific factors in credit scores or credit based insurance scores results in "negative or differential treatment" of ECOA protected classes?

6. What is an appropriate methodology to determine whether there are factors that are not considered by credit scores or credit based insurance scores that result in "negative or differential treatment" of ECOA protected classes?

7. In order to address paragraphs (a)(2) and (a)(3) of Section 215, data are needed on the geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, or creed of borrowers, potential borrowers, insurance customers, or

potential insurance customers. Are these data available, and if so, where?

8. If the data discussed in question 7 are not available, what proxies are available for the geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, or creed of borrowers, potential borrowers, insurance customers, or potential insurance customers?

9. If there are proxies for the geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, or creed of borrowers, potential borrowers, insurance customers, or potential insurance customers, what type of analysis would allow inferences to be drawn using the proxies instead of actual data on individual characteristics? What limitations are there to the inferences that can be drawn using proxies in place of data on individual characteristics?

10. One potential proxy for individual characteristics may be Census data about the location where a borrower or insurance customer resides. What type of analysis would allow inferences to be drawn using data about the characteristics of the location where a borrower or insurance customer resides instead of data on individual characteristics? What limitations are there to the inferences that can be drawn using data about the characteristics of the location where a borrower or insurance customer resides in place of data on individual characteristics?

Authority: Sec. 112(b), Pub. L. 108-159, 117 Stat. 1956 (15 U.S.C. 1681c-1).

By direction of the Commission.

Donald S. Clark

Secretary

[Billing Code 6750-01-P]

FEDERAL TRADE COMMISSION
RIN [3084-AA94]

Public Comment on Data, Studies, or Other Evidence Related to the Effects of Credit Scores and Credit-Based Insurance Scores on the Availability and Affordability of Financial Products

AGENCY: Federal Trade Commission
ACTION: Notice and request for public comment.

SUMMARY: The Fair and Accurate Credit Transactions Act of 2003 ("FACT Act" or "Act") requires the Federal Trade Commission ("FTC" or "Commission") and the Federal Reserve Board ("Board") to conduct a study on the effects of credit scores and credit-based insurance scores on the availability and affordability of financial products. These products include credit cards, mortgages, auto loans, and property and casualty insurance. As part of its efforts to fulfill its obligations under the Act, the FTC seeks public comment on any evidence the FTC and the Board should consider in conducting the study.

DATES: Comments must be received by April 25, 2005.

ADDRESSES: Public comments are invited, and may be filed with the Commission in either paper or electronic form. Comments filed in paper form should refer to "FACT Act Scores Study" both in the text and on the envelope, to facilitate their organization, and should be mailed or delivered to: Federal Trade Commission/Office of the Secretary, Room H-159 (Annex Z), 600 Pennsylvania Avenue, N.W., Washington, D.C. 20580. The FTC requests that any comment filed in paper form be sent by courier or overnight service, if possible, because U.S. postal mail in the Washington area and at the Commission is subject to delay due to heightened security precautions. Comments may be filed in electronic form by clicking on the following: https://secure.commentworks.com/FTCCreditScoreStudy/ and following the instructions on the web-based form. If a comment contains confidential information, it must be filed in paper (rather than electronic) form, and the first page of the document must be clearly labeled "Confidential."[144]

To ensure that the Commission considers an electronic comment, you must file it on the web-based form at https://secure.commentworks.com/FTCCreditScoreStudy/. You also may visit http://www.regulations.gov to read this Notice, and may file an electronic comment through that website. The Commission will consider all comments that regulations.gov forwards to it.

The FTC Act and other laws the Commission administers permit the collection of public comments to consider and use in this proceeding as appropriate. All timely and responsive public comments, whether filed in paper or electronic form, will be considered by the Commission, and will be available to the public on the FTC Web site, to the extent practicable, at www.ftc.gov. As a matter of discretion, the FTC makes every effort to remove home contact information for individuals from the public comments it receives before placing those comments on the FTC Web site. More Information, including routine uses permitted by the Privacy Act, may be found in the FTC's privacy policy, at http://www.ftc.gov/ftc/privacy.htm.

FOR FURTHER INFORMATION CONTACT:

---

[144] Commission Rule 4.2(d), 16 CFR 4.2(d). The comment must also be accompanied by an explicit request for confidential treatment, including the factual and legal basis for the request, and must identify the specific portions of the comment to be withheld from the public record. The request will be granted or denied by the Commission's General Counsel, consistent with applicable law and the public interest. See Commission Rule 4.9(c), 16 CFR 4.9(c).

Jesse Leary, Deputy Assistant Director, (202) 326-3480, Division of Consumer Protection, Bureau of Economics, Federal Trade Commission, 600 Pennsylvania Avenue, N.W., Washington, DC 20580.

SUPPLEMENTARY INFORMATION:

I. Background

The FACT Act was signed into law on December 4, 2003. Fair and Accurate Credit Transactions Act of 2003, Pub. L. No. 108-159 (2003). In general, the Act amends the Fair Credit Reporting Act ("FCRA") to enhance the accuracy of consumer reports and to allow consumers to exercise greater control regarding the type and amount of marketing solicitations they receive. The Act contains a number of provisions intended to combat consumer fraud and related crimes, including identity theft, and to assist its victims. Finally, the Act requires that a number of studies be conducted on credit reporting and related issues.

Section 215 of the FACT Act requires the FTC and the Board, in consultation with the Office of Fair Housing and Equal Opportunity of the Department of Housing and Urban Development, to conduct a study on the effects of credit scores and credit based insurance scores on the availability and affordability of financial products. These products include mortgages, auto loans, credit cards, and property and casualty insurance. Section 215 further requires the FTC and the Board to study: 1) "the statistical relationship, utilizing a multivariate analysis that controls for prohibited factors under the Equal Credit Opportunity Act and other known risk factors, between credit scores and credit-based insurance scores and the quantifiable risks and actual losses;" and 2) "the extent to which, if any, the use of credit scoring models, credit scores, and credit-based insurance scores impact on the availability and affordability of credit to the extent information is currently available or is available through proxies, by geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, and creed, including the extent to which the consideration or lack of consideration of certain factors by credit scoring systems could result in negative or differential treatment of the protected classes, under the Equal Credit Opportunity Act, and the extent to which, if any, the use of underwriting systems relying on these models could achieve comparable results through the use of factors with less negative impact."

The study is due on December 4, 2005.

II. Request for Comments

The Act requires the FTC to seek public input about "the prescribed methodology and research design of the study." As part of its efforts to fulfill its obligations under the Act, the FTC, (in a *Federal Register* notice dated June 18, 2004, see 69 FR 34167) sought public comment on methodological aspects of the study. The FTC received comments in response to that notice, and the FTC and the Board are considering them as they conduct the study. In the present request, the FTC seeks comment on specific studies, data, or other evidence that might be useful for the study. Although we enumerate a set of questions below, we encourage commenters to provide information on any aspects of credit scores, credit-based insurance scores, and the effects of scores on the relevant markets that would be useful to the study. In particular, the FTC seeks information that bears on the following questions:

A. Credit Scores and Credit:

1. Specifically, how are credit scoring models developed? Who develops credit

scoring models? What data and methodologies are used to develop credit scoring models? What factors are used in credit scoring models? Why are those factors used?
What other factors have been considered for use in credit scoring models, but are not used? Why are those other factors not used? Are there benefits or disadvantages, either to creditors or consumers, from the use of particular factors by credit scoring models?

2. How many different credit scoring models are in use today? What different types of general purpose or specialized credit scoring models are available?
Who offers credit scores?

3. How are credit scores used? Who uses credit scores, and how widely are they used? How do they fit into the underwriting process for mortgages, auto loans, credit cards, and other credit products? For what purposes are credit scores used, other than the initial underwriting or pricing decision?

4. How has the use of credit scores changed over time? When were they first used for each type of financial product (credit cards, mortgages, auto loans, etc.)? How has their use expanded to encompass different groups of borrowers (*e.g.*, lower income borrowers, urban/rural borrowers, borrowers with poor credit histories, borrowers with non-traditional credit histories)? If the use of credit scores has expanded to encompass Different groups of borrowers, how has this affected the price or availability of credit to those borrowers?

5. Has the use of credit scores affected the price and availability of mortgages, auto loans, credit cards, or other credit products? If so, are there estimates of the type and size of such changes? Have some groups of consumers experienced cost reductions while others have experienced cost increases? Have some groups of consumers experienced

greater access to credit while others have experienced reduced access?

6. Has the use of credit scores affected the amount of credit made available to consumers? Has it affected initial loan-to-value ratios at which auto loans or mortgages (first- or second-lien) are originated to different groups of borrowers? Has it affected credit limits on credit cards and home equity lines of credit for different groups of borrowers?

7. How has the use of credit scores affected the costs of underwriting and/or the time needed to underwrite?

8. What impact has the use of credit scores had on the accuracy of underwriting decisions? What impact has the use of credit scores had on the share of applicants that are approved for mortgages, auto loans, credit cards, or other credit products? What impact has the use of credit scores had on the default rates of mortgages, auto loans, credit cards, or other credit products? Have the sizes of such changes or effects been estimated and reported?

9. Has the use of credit scores affected the cost and availability of credit to consumers with poor credit histories? If so, how? What effect has it had on the use of credit by consumers with poor credit histories?

10. How has the use of credit scores affected the cost and availability of credit to consumers with no credit history? What effect has it had on the use of credit by consumers with no credit history?

11. How has the use of credit scores affected refinancing behavior for mortgage, auto, or student loans? How has it affected the average life of revolving lines of credit (including credit cards)?

12. Has the use of credit scores and credit scoring models impacted the availability or cost of credit to consumers by geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, or creed? If so, how has it impacted each such category? What are the estimated sizes of any such changes for each of the above categories?

13. To what extent does consideration or lack of consideration of certain factors by credit scoring systems result in negative or differential treatment of those categories of consumers who are protected under the Equal Credit Opportunity Act ("ECOA") (*e.g.*, race, color, religion, national origin, sex, age, and marital status)?

14. To what extent, if any, could the use of underwriting systems that rely on scoring models achieve comparable results through the use of factors with less negative impact on those categories of consumers who are protected under the ECOA?

15. What steps, if any, do score developers, lenders, or other users of credit scores take to ensure that the use of credit scores does not result in negative or differential treatment of protected categories of consumers under the ECOA? Have score developers, lenders, or other users of credit scores changed the way credit scores are developed or used in order to avoid negative or differential treatment of protected categories of consumers under the ECOA? Are any particular credit history factors not used because of actual or potential negative or differential treatment of protected categories of consumers under the ECOA? If so, what are they?

16. Has the use of credit scores caused a change in the rate of home ownership? What is the estimated size of such a change?

17. Has the use of credit scores caused a change in the method and amount of pre-

screening consumers for credit offers? What effects has this had on the terms offered to consumers?

18. What specific role do credit scores play in granting “instant credit?” What impact have credit scores had on the availability and use of instant credit?

19. How has the use of credit scores affected companies' ability to enter new lines of business or expand activities in the various credit industries?

20. What role does credit scoring play in secondary market activities? In what ways has the availability of credit scores affected the development of the secondary market for credit products? Has the use of credit scoring increased or decreased creditors’ access to capital? In what ways?

21. How are credit scores used to manage existing credit accounts, such as credit card accounts? How has the use of credit scores affected the way credit accounts are managed? How are credit scores used in the servicing of mortgages, and how has the use of credit scores affected the way mortgages are serviced?

22. How are records of inquiries used by credit scoring systems? Does concern about the possible effects on their credit scores affect consumers’ credit shopping behavior? If so, what impact does this have on the consumers or on competition in the various credit markets?

23. How does the use of credit scores affect consumers with inaccurate information on their credit reports? How does the use of credit scores affect consumers who have been the victims of identity theft?

24. Are there particular forms of inaccuracy or incompleteness in the credit reporting system, such as incomplete reporting by creditors, that affect either the

usefulness of credit scores to lenders or the benefits or disadvantages of scoring to consumers? What are those types of inaccuracies or incompleteness? How do they affect the usefulness of credit scores to lenders or the benefits or disadvantages of scoring to consumers?

B. Credit-Based Insurance Scores and Property and Casualty Insurance:

1. Specifically, how are credit-based insurance scoring models developed? Who develops credit-based insurance scoring models? What data and methodologies are used to develop credit-based insurance scoring models? What factors are used in credit based insurance scoring models? Why are those factors used? What other factors have been considered for use in credit-based insurance scoring models, but are not used? Why are those other factors not used? Are there benefits or disadvantages, either to insurers or consumers, from the use of particular factors by credit-based insurance scoring models?

2. How many different credit-based insurance scoring models are in use today? Who offers credit-based insurance scores?

3. How are credit-based insurance scores used? Who uses credit-based insurance scores, and how widely are they used? How do they fit into the underwriting and rating process for automobile and homeowners insurance?

4. Has the use of credit-based insurance scores affected the price and availability of automobile and homeowners insurance? We are especially interested in evidence containing estimates of the size of such changes. Have some groups of consumers experienced cost reductions while others have experienced cost increases? If so, which consumers have experienced reductions and which have experienced increases, and what are the magnitudes of those changes? Have some consumers experienced dramatic

increases in their insurance premiums, solely as the result of the introduction of credit-based insurance scoring? If so, what has been the impact of this rise in premiums on these consumers?

5. How has the use of credit-based insurance scores affected the costs of underwriting and rating and/or the time needed to underwrite and rate?

6. How has the use of credit-based insurance scores affected the accuracy of underwriting and rating decisions? Have the sizes of such changes been estimated and reported?

7. Has the use of credit-based insurance scores affected the amount of automobile or homeowners insurance purchased by consumers? Has it affected the limits or deductibles that consumers select when purchasing automobile or homeowners insurance? Has it affected the number of drivers who drive without insurance? Has it affected the number of homeowners that have no homeowners insurance? What are the estimated sizes of such changes?

8. How has the use of credit-based insurance scores affected the cost and availability of automobile or homeowners insurance to consumers with poor credit histories? What effect has it had on the purchasing of automobile or homeowners insurance by consumers with poor credit histories?

9. Has the use of credit-based insurance scores affected the cost and availability of automobile or homeowners insurance to consumers with no credit history? If so, how? What effect has it had on the purchasing of automobile or homeowners insurance by consumers with no credit histories?

10. How has the use of credit-based insurance scores impacted the availability or

cost of insurance to consumers by geography, income, ethnicity, race, color, religion, national origin, age, sex, marital status, or creed? What are the estimated sizes of such changes for each of the above categories?

11. To what extent does consideration or lack of consideration of certain factors by credit-based insurance scoring systems result in negative or differential treatment of protected classes of consumers, that is, the same categories of consumers against whom discrimination is prohibited under the ECOA (*e.g.* race, color, religion, national origin, sex, age, and marital status)?

12. To what extent, if any, could the use of underwriting systems relying on credit-based insurance scoring models achieve comparable results through the use of factors with less negative impact on consumers in the ECOA protected categories?

13. What steps, if any, do score developers or insurance companies take to ensure that the use of credit-based insurance scores does not result in negative or differential treatment of protected categories of consumers listed in the ECOA? Have score developers or insurance companies changed the way credit-based insurance scores are developed or used in order to avoid negative or differential treatment of protected categories of consumers listed in the ECOA? Are any particular credit history factors not used because of actual or potential negative or differential treatment of protected categories of consumers listed in the ECOA? If so, what are they?

14. Has the use of credit-based insurance scores caused a change in the method and amount of pre-screening consumers for insurance offers? What effects has this had on the terms offered to consumers?

15. How has the use of credit-based insurance scores affected companies' ability

to enter new lines of the automobile or home- owners insurance business?

16. If the use of credit-based insurance scores has affected the costs individual consumers pay for insurance, has it (i) caused a change in the overall average cost of insurance for consumers?; (ii) changed the distribution of individual costs?; or (iii) Caused any other change in the costs to consumers? What are the magnitudes of any such changes?

17. Would an analysis of the share or number of consumers that purchase automobile or homeowners insurance from “involuntary,” “pooled risk,” “assigned risk,” or other types of insurance other than insurance offered on a voluntary basis by private insurers, be informative about the price and/or availability of automobile or homeowners insurance? Would an analysis of the share of drivers that drive without automobile insurance be informative about the price and/or availability of automobile insurance?

18. What impact, if any, does banning or limiting the use of particular underwriting or rating factors, such as gender, territory, or credit-based insurance score, have on the price or availability of automobile or homeowners insurance? Has the prohibition on the use of credit-based scores for insurance in particular states had any impact on the price or availability of automobile or homeowners insurance for consumers in those states? If so, what has that impact been? If the use of credit-based insurance scores was not allowed in additional states, what impact would this have on the price or availability of automobile or homeowners insurance? Are there, or would there be, any specific effects on those insurance consumers who are within protected categories listed in the ECOA?

19. How are records of inquiries used by credit-based insurance scoring systems?

Does concern about the possible effects on their credit-based insurance scores affect consumers' insurance-shopping behavior? If so, what impact does this have on competition in the insurance markets?

20. How does the use of credit-based insurance scores affect consumers with inaccurate information on their credit reports? How does the use of credit-based insurance scores affect consumers who have been the victims of identity theft?

21. Are there particular forms of inaccuracy or incompleteness in the credit reporting system, such as incomplete reporting by creditors, that affect either the usefulness of credit-based insurance scores to insurers or the benefits or disadvantages of scoring to consumers? What are those types of inaccuracies or incompleteness? How do they affect the usefulness of credit-based insurance scores to insurers or the benefits or disadvantages of scoring to consumers?

Authority: Sec. 112(b), Pub. L. 108-159, 117 Stat. 1956 (15 U.S.C. 1681c-1).

By direction of the Commission.

Donald S. Clark

Secretary

APPENDIX C

THE AUTOMOBILE POLICY DATABASE

APPENDIX C. The Automobile Policy Database

The FTC constructed the database of automobile policies used to do the analysis for this report by combining policy data from five large auto insurance firms submitted with data from a range of additional sources. This Appendix describes that process.

C.1. The EPIC Database

The automobile policy data in the FTC database were originally collected for a study conducted by EPIC, a firm of consulting actuaries, that was released in 2003.[145] The EPIC database was constructed by randomly sampling from the policies in place at the participating firms between July 1, 2000 and June 30, 2001. Data on policies that were in place throughout the sample year were collected for the entire year. Data on policies of customers that left a firm during the year were collected until the policy ended, and data on the policies of customers that joined were collected from the date the policy began until the end of the year. While the EPIC report did not include information on the number of cars in their database, it did provide information on the total "earned car years." An "earned car year" is equivalent to one year of insurance coverage for one car. The EPIC database contained roughly 2.7 million earned car years.

The sampling of policies was done in a way that produced roughly the same number of records from each firm. This means that the larger firms in the database are under-represented, relative to their market share. All cars covered by a sampled policy were included in the sample. The samples were drawn to ensure that some minimum number of policies would be available for each state. This means that drivers in small states were over-represented in the sample.[146]

---

[145] Michael J. Miller and Richard A. Smith, *The Relationship of Credit-Based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity: An Actuarial Study by EPIC Actuaries, LLC* (June 2003) [hereinafter EPIC Study], *available at* http://www.progressive.com/shop/EPIC_CreditScores.pdf.

EPIC received data on the cars and drivers covered by each policy. Car information included vehicle identification number (VIN), miles driven, coverages, limits, deductibles, premiums, and claims paid. Driver information included most standard risk variables, including age, gender, marital status, and driving history (*e.g.*, violations). Important risk variables missing from the data were prior claims (on accidents at companies other than the customer's current company) and territory. EPIC did attempt to control for territory in their analysis by using the population density of each ZIP code, based on Census data.

Claims were included in the data if they were for events that occurred between July 1, 2000 and June 30, 2001. The samples were drawn in the second half of 2002, and information on claims is as of June 30, 2002. For some claims, especially bodily injury liability claims, the reported amount paid out on the claim may not reflect the actual ultimate cost of the claim. This is because the process of determining the final cost of a claim can take a very long time, especially if the claim goes to litigation. For claims that were not yet resolved, any reserves for the claims were included as an amount paid.

Credit-based insurance scores had never been calculated for many of the policies in the database. For those that had been scored, different companies may have used different models, and the models may have varied by state. The credit scores EPIC obtained for the study were ChoicePoint Attract Standard Auto scores. Scores were only calculated for one person, the first named insured, for each policy. This means that the same score was assigned to each car covered by a policy, even if a different person was the primary driver of that car.[147] Credit history data used by ChoicePoint to calculate scores came from the June 2000 archives of Experian (just before the beginning of the sample period). There were three possible outcomes for each individual submitted for scoring: a score, a "no-hit," meaning a credit report for the person could not be located in Experian's records, and a "thin-file," meaning a credit report for the person could be located, but it did not contain enough information to calculate a score.

---

[146] All of the analysis presented in the body of the report uses data that have been weighted to be geographically representative.

High-risk drivers are likely under-represented in the database. None of the firms provided data on "residual market" policies. These are policies purchased through state-run plans that offer access to insurance for customers who are unable to purchase insurance in the normal "voluntary" market. They make up less than 2% of the total market for automobile insurance. In addition, while four of the five firms that submitted data to the FTC did sell policies to high-risk drivers, two of them did so through subsidiaries that did not use the same data systems, and therefore policies from the high-risk subsidiaries were not included in the sample. These subsidiaries represented less than 5% of the total business of any one firm, and less than 2% of the total business of the five firms. Although these are small portions of these firms' total customers, it is quite possible that the sample under-represents the highest-risk portion of the insurance market. For this reason, we conducted an analysis that focused on the highest-risk portion of the sample that was collected. This analysis is described in Appendix F.

C.2 The FTC Database

The database analyzed by the FTC is a subset of the original EPIC database. Not all of the firms that contributed data to the EPIC database agreed to have their data forwarded to the FTC for this study. Data from five firms were submitted to the FTC. These five firms together represented 27% of the U.S. market of automobile insurance in 2000 (the time period covered by the data).

---

[147] This is a form of measurement error that should have the effect of understating the relationship between credit score and claims.

The database submitted by the five firms includes over 2.5 million records. Each record has data on one car for up to one year. Many records cover only part of the year, either because customers commenced or discontinued coverage during the year, or because the company generated a separate record each time a policy was renewed or modified. Adjusting for the period of time covered by each record, the total number of "car-years" in the database is just over 1.8 million. Many of the policies in the database cover more than one car; the total number of policies in the database is 1.4 million.

The FTC combined the information the insurance firms submitted with data from a number of other sources. The agency obtained additional information to broaden the range of credit history variables analyzed; to improve the set of other risk controls in the analysis; to provide an independent measure of claims; and to analyze issues relating to race, ethnicity, income, and national origin. In constructing the database, the FTC never took possession of any personally identifying information. The following describes the data that were collected and the process by which they were collected.

C.2.1 Additional Information Obtained for the Full Sample

Core Policy Data and ChoicePoint Credit Scores

The participating firms submitted their samples of policy data to EPIC.[148] EPIC forwarded the data to ChoicePoint. ChoicePoint calculated and appended the Attract

---

[148] During the course of this project, EPIC was purchased by Tillinghast/Towers Perrin Consulting. For simplicity, we refer to "EPIC" throughout this appendix, even though some of the steps in the data collection and preparation process took place after the change in ownership.

Standard Auto credit-based insurance scores, stripped off the names and addresses, and created a new anonymous unique identifier. ChoicePoint then returned the database to EPIC.

EPIC standardized the coding of the data and combined the data from the five firms into a single database. When a particular variable was always missing for a particular company, a small portion (5%) of records of that variable for other companies were chosen at random and changed to missing. This was done to mask which policies came from the same company. The combined database was then forwarded to the FTC.

Territorial Risk Variable

The five firms also submitted to EPIC data on earned car years and claims on property damage liability policies by ZIP code for a three-year period from 2000 to 2002, for their full book of business. EPIC combined the data from the five firms to calculate ZIP-code level average property damage liability pure premiums (*i.e.*, average dollars paid out per year of coverage per car).[149] This is an improvement over the original Census-based population density measure that EPIC used in its report. The new ZIP code risk variable was included in the policy database EPIC forwarded to the FTC.

Geographic Location Information and Census Data

ChoicePoint used commercial mapping software to match the addresses of the drivers in the database to Census location information (a process commonly referred to as "geo-coding"). These data were sent to EPIC, and forwarded to the FTC with the core policy database. ChoicePoint was able to determine the Census block location for 95% of the overall sample, and 98% of the sub-sample for which Social Security

---

[149] For ZIP codes with fewer than 3,000 property damage liability claims, data from surrounding ZIP codes were also used to calculate average pure premiums.

Administration race and ethnicity data were obtained (see below for a discussion of the Social Security Administration data). FTC staff used the Census location information to append data on race, ethnicity, vehicle ownership, and income from the 2000 Census.

ChoicePoint Credit History Variables

In the process of calculating the ChoicePoint credit scores, ChoicePoint generated and maintained 180 credit history variables for each person for whom Experian was able to locate a credit report. These are a set of variables that ChoicePoint has developed over time for its score-building research that are intended to capture all important information contained in a credit report. These 180 credit history variables are from the June 2000 Experian credit report archive. ChoicePoint forwarded the credit history variables directly to the FTC.

CLUE Data

ChoicePoint collects data on claims from most major automobile insurance firms in the United States. These data allow firms to determine whether a potential new customer has filed a claim under a previous policy with another firm, and use that information in underwriting and rating. The database is referred to as the Comprehensive Loss Underwriting Exchange ("CLUE").

Pursuant to two 6(b) orders, the FTC obtained the CLUE records for everyone in our database for the period July 1995 – June 2003:[150] five years prior to the year covered by the firm-submitted data, the year covered by the firm data, and two years after.

---

[150] The CLUE database maintains records on individual claims, with name and address and other identifying information about the policy on which the claim was filed. The CLUE records that the FTC obtained were found by matching the names and addresses in the company-submitted data to the CLUE database. Claims, therefore, were only located for people who had the same address in the company data and the CLUE database, and the claims of people who had moved were not located.

ChoicePoint sent the CLUE data directly to the FTC.

Hispanic Surname Match

ChoicePoint forwarded to Experian a database containing the names and addresses of the individuals in the sample, along with the anonymous unique identifier created by ChoicePoint. The FTC forwarded to Experian a file containing a list of Hispanic surnames created by the Census Department following the 1990 Census.[151] Experian matched the last names of all of the drivers in the database against the list of Hispanic surnames. Experian then forwarded directly to the FTC a database containing only the anonymous unique identifier for each record in the database, and an indicator for whether the surname of the person associated with that record was on the Census list of Hispanic surnames.

Vehicle Characteristics

Included with the database EPIC forwarded to the FTC was a 10-digit Vehicle Identification Number (VIN). These are not enough digits to identify a particular vehicle, but enough to identify make and model. The 10-digit VINs were matched to Edmunds data on a range of vehicle characteristics, including vehicle body type (*e.g.*, sedan, pickup truck, etc.), engine displacement, and safety features.

C.2.2 Additional Information Obtained for a Sub-Sample of 400,000

Some data were obtained for only a sub-sample of the records. A sub-sample was used for budgetary reasons. The sub-sample consisted of 400,000 of the 1.4 million policies in the FTC database. Using a smaller sample can reduce the power of statistical tests. To minimize that effect, the sub-sample was drawn using stratification: all policies with claims were included in the sub-sample, and policies without claims were sampled at a rate sufficient to bring the total to 400,000.[152] This results in a much smaller reduction in statistical power than simple, un-stratified random sampling. ChoicePoint conducted the sampling following directions from the FTC.

---

[151] The list and a paper that describes how it was developed are available at: http://www.census.gov/population/documentation/twpno13.pdf

FICO Scores

ChoicePoint arranged for Experian to match the names and addresses of the first named insureds of the 400,000 policy sub-sample against the June 2000 credit history archive, and calculate a FICO "Standard Auto, Greater than Minimum Limits" credit-based insurance score. Experian forwarded the FICO scores (or an indicator for why a score could not be calculated – either "no-hit" or "thin file") directly to the FTC.

SSA Data on Race, Ethnicity, National Origin, and Gender

Whenever someone applies for a Social Security card, the Social Security Administration (SSA) attempts to collect information on race, ethnicity, national origin, and gender. That information is recorded in the SSA's "Numident" file. Experian attempted to locate Social Security Numbers (SSNs) and dates of birth (DOBs) for the 400,000-person sub-sample in Experian's consumer credit history files. DOBs were only used when an actual day, month, and year could be found. Experian located an SSN or valid DOB for 324,563 individuals. The name, SSN, DOB, and the anonymous identifier for those individuals were forwarded to the SSA. The SSA matched name, SSN, and DOB against the Numident file, and was able to locate information for 308,746 individuals. The SSA then deleted the names, SSNs, and DOBs, and forwarded to the

---

[152] Of the 400,000, 56% had a claim in at least one coverage, and 44% had no claim. We used the sampling probabilities to construct sampling weights, which are used throughout the analysis to keep the sub-sample representative of the overall sample.

FTC the anonymous unique identifier and data on race, ethnicity, national origin, and gender.

APPENDIX D

MODELING AND ANALYSIS DETAILS

APPENDIX D Analysis and Modeling Details

D.1 Intermediate Analysis and Data Preparation

The process of preparing and analyzing the FTC database included several intermediate analyses and data preparation steps that require further explanation. First, the race and ethnicity data in the database were from several imperfect sources, and were combined in a way to take advantage of the strengths of each. Second, the sample likely was not representative of the national population of automobile insurance customers, and so was weighted to be representative by geography, and race and ethnicity. Finally, the risk models were not run on the full sample, mainly because race and ethnicity data are only present for a sub-set of the policies. The process of creating the modeling sample is described below.

D.1.1 Using Race and Ethnicity Data

The data on race and ethnicity in the FTC database come from three sources: SSA data, a Hispanic surname match, and Census information about the racial and ethnic makeup of the location where each individual lives. The SSA data have the two most important attributes of race/ethnicity data: they are at the individual level, and they are self-reported. The Hispanic surname match is at the individual level, but is not self-reported. (Comparing the SSA data and the Hispanic surname match shows that there are many people who have a Hispanic surname who do not report themselves to be Hispanic, and vice-versa.) The Census data come from self-reports, but they are only available for geographic areas, not for individuals.

The SSA data do have an important limitation. Prior to 1981, the only available answers to the race/ethnicity question were: “White,” “Black,” or “Other.” After 1981, the choices were expanded to include "Hispanic," "Asian, Asian-American, or Pacific Islander," and "North American Indian or Native Alaskan," and the "White" and "Black" categories were specifically labeled "non-Hispanic."[153] The "Other" option was dropped. Our only option for identifying Hispanics, Asians, and Native Americans among people for whom we only had pre-1981 responses was to make inferences using the information we did have.

The SSA was able to locate the records of 308,746 people, out of the 324,563 for whom Experian was able to locate an SSN or a valid date of birth. Of those, 10,661 did not have a valid response to a race/ethnicity question. Of the 298,085 people for whom we had valid race/ethnicity data, 162,755 had only a pre-1981 response. These are the people for whom we only had answers for the limited race/ethnicity options. We did, however, have pre- *and* post-1981 responses for 91,519 people. This allows us to evaluate how people identified themselves when given the limited set of race/ethnicity choices, and how they subsequently identified themselves when given the broader set of choices. Based on those patterns, we determined that very few people who answered "Black" pre-1981 chose some other option post-1981, and very few people who answered "White" pre-1981 chose "Black" post-1981. For this reason, anyone who answered "Black" pre-1981 was identified as African American, and no one who answered "White" pre-1981 was identified as African American.

The remaining challenge was to try to determine how someone who answered

---

[153] The post-1981 options raise other concerns. In particular, "Hispanic" is presented as a mutually exclusive alternative to the other options. In recent Census questionnaires, "Hispanic/Non-Hispanic" information is collected separately from race information. In our data, we find a lower number of people with Hispanic surnames self-identifying as Hispanic, post-1981, than does the Census. This is likely due to the fact that the Census questionnaire, unlike the SSA questionnaire, collects race and ethnicity data separately.

"White" or "Other" pre-1981 would have answered if given the broader post-1981 set of choices. We did that using a statistical analysis of individuals for whom we have pre- and post-1981 responses. The analysis was based on the following factors: the pre-1981 response ("White" or "Other"); whether someone had a Hispanic surname (from the surname match); country of birth (from the SSA data); gender (from the SSA data), and the racial/ethnic makeup of the Census block where the person lived.

We split the group of people who have both a pre-1981 and a post-1981 SSA race/ethnicity answer into cells using the following characteristics:

- Pre-1981 SSA race/ethnicity answer (*i.e.*, "white" or "other") (2 categories)
- Gender (2 categories)
- Region of Birth, based on Country of Birth from SSA data (4 categories):
  - U.S. born.
  - "Hispanic" Countries: Countries of birth where more than half of the people born in that country identified themselves as Hispanic in their post-1981 SSA race/ethnicity response.
  - "Asian" Countries: Countries of birth where more than half of the people born in that country identified themselves as Asian in their post-1981 SSA race/ethnicity response.
  - All Other Countries: Countries of birth that were not included in the three prior categories (these are mainly countries in Europe, the Middle East, and Africa).
- Hispanic surname match flag (2 categories)

This generated 32 cells (*i.e.*, 2x2x4x2). Within each cell, we ran a simple logit model to predict the probabilities that someone would answer "Hispanic" vs. "white", "Asian" vs. "white", or "Black" vs. "white" (the latter only for people who answered "Other" pre-1981) using the relative Census block race/ethnicity concentration for that race/ethnic group vs. non-Hispanic whites as the explanatory variable.[154]

We then imputed the probability of being of each race for the individuals in each cell for whom we only have a pre-1981 race/ethnicity answer. This was a two-step process. We first estimated the probability of being of a given race/ethnicity relative to the probability of being non-Hispanic white, and then used a log-odds ratio calculation to determine the probability of being of a given race or ethnicity.[155]

To use the predicted probabilities that come out of this process, we generated a record for each race/ethnicity that was estimated to have a positive probability for each person. Each of these records was identical, except for the race/ethnicity variable. We included the multiple records in the analysis, giving each record a weight equal to the predicted probability. For example, someone who is predicted to be non-Hispanic white with 85% probability, Asian with 10% probability, and Hispanic with 5% probability will have three records in the database. One record will have "non-Hispanic white" as the race/ethnicity, and a weight of .85; one record will have "Asian" as the race/ethnicity, and a weight of .1; and one record will have "Hispanic" as the race/ethnicity, and a weight of .05.[156]

---

[154] For several cells where everyone, or nearly everyone, gave the same post-1981 answer we simply assigned everyone in that cell to that category with probability one. For example, all men who answered "other" pre-1981, were born in a Hispanic country, and had a Hispanic surname were considered to be Hispanic.

[155] The predicted values from a logit are bounded between zero and one, and therefore a logit model gives every person for whom we predicted race/ethnicity a positive predicted probability of being each race or ethnicity. This is true even for people who lived on blocks with no residents of that race or ethnicity, according to the 2000 Census. We therefore reset the predicted probabilities of being of a given race or ethnicity to zero if someone lived on a block with no residents of that race or ethnicity and had a predicted probability from the logit model of being of that race or ethnicity that was less than 1%. As discussed in Appendix F, we also ran the analysis without that restriction and the results were unaffected.

[156] We also estimated the probability of being Native American, but there were so few Native Americans in the sample that we did not include them in the analysis.

D.1.2 Nationally Representative Weighting

One limitation of the database is that it was a random sample of policies of customers of five insurance firms, not a random sample of all insurance customers in the nation. We did not have sufficient information about the automobile insurance market as a whole to know exactly how well our sample represented the entire market.[157] Because much of the analysis presented in this report focuses on the relationship between race, ethnicity, income and credit history and insurance risk, the racial, ethnic, and income mix of the sample could have affected the results.

We did not know the racial, ethnic, and neighborhood-income makeup of car insurance customers nationwide. We did, however, observe the racial, ethnic, and income breakdowns of car ownership, using the 2000 Census. This is shown in column (a) of Table A.1.[158] Column (b) shows the same breakdowns in our sample.[159] Comparing our sample with Census data on car owners, we see that our sample under-represented minorities and residents of low-to-moderate income tracts, and over-represented non-Hispanic whites and residents of upper-income tracts. We did not know how much of this difference is due to differences between the customers of the companies in the sample relative to the market as a whole, versus differences between the racial, ethnic and income make-up of the general population of car owners relative to the population of car owners with insurance.

---

[157] We do know that the sample likely under-represents the highest-risk portion of the market. As described in Appendix F, the robustness checks appendix, we also estimated risk models for the riskiest segment of the sample.

[158] The distribution of race and ethnicity for vehicle owners in the overall Census data, which was used as the "target" for the weighting, was adjusted using Census race and ethnicity data for the full sample of 1.4 million policies and the sub-sample for which we obtained SSA race and ethnicity data. This was done so that if the weights developed on the sub-sample were applied to the full sample (which would require obtaining SSA race and ethnicity data for the full sample), that full sample would have the correct distribution of race and ethnicity.

[159] The racial and ethnic makeup of the FTC sample is based on the SSA race and ethnicity data, including the imputed results for people for whom we only have pre-1981 data.

To make our sample close to nationally representative, we weighted the sample using a two-step process. We first created a geographic weight at the Census tract level. Our database contained cars from most Census tracts in the country. There were 64,946 tracts with cars in the 2000 Census, and our database contains records from 62,964 of those tracts.[160] We therefore could make our sample almost perfectly geographically representative of the entire country by applying a weight that was the ratio of the share of all cars in the country that are in a tract over the share of cars in our sample that are in that tract.[161] Column (c) of Table A.1 shows the racial, ethnic and income breakdown after weighting the sample in this way. The weighted sample was now almost perfectly nationally representative by income group, because income is measured at the tract level, but minorities were still under-represented. We therefore applied a second weight, which was the ratio of the share of cars owned by each racial or ethnic group in the country over the share of cars owned by each racial or ethnic group in the sample after applying the tract weights. Column (d) of Table A.1 shows the racial, ethnic, and income breakdowns after applying those weights.[162] The racial and ethnic proportions were now the same as those for the nation as a whole, by construction. Adding this second weight did make the weighted sample slightly over-representative of residents of low-to-moderate income tracts, but it was very close to the national numbers. We used these weights throughout the analysis, except where noted.

---

[160] The 62,964 tracts are in the full database. Tract weights are applied to the full database, and the second step – the race-weight step – is done with the sub-sample for whom we have SSA race and ethnicity data.

[161] To be precise, the measure in the FTC database is the share of property damage liability earned car years by tract. There were a small number of tracts with very small number of earned car years (for example, someone may have only had a week of coverage) that resulted in very large tract weights. We capped the tract weights at the 99.95 percentile of their distribution. Even with that cap, there were some outliers once claims paid were adjusted for earned car years. Removing these outliers did not affect the results of the analysis. These results are discussed in Appendix F, the robustness check appendix.

[162] Because individuals with imputed race and ethnicity are represented by multiple records in the database, each record received the appropriate nationally representative weight associated with the race or ethnicity of that record.

CLUE data were analyzed using the full sample of 1.4 million policies. Because the main weights were developed to apply to SSA race and ethnicity data, and we only have SSA race and ethnicity data for a sub-sample, we cannot generate those weights for the full sample of 1.4 million. Instead, we first developed a set of weights to make the sample geographically representative at the tract level, and then calculated race and ethnicity weights based on Census block-level race data.

D.1.3 The Modeling Sample

Most of the analysis presented in the report was conducted using a sub-sample of the original database. As discussed in Appendix C, which describes the construction of the database, the FTC only obtained SSA race and ethnicity data for a stratified sub-sample of the database. Although not all of the analysis required the use of race/ethnicity data, the sub-sample with that information was used throughout the report for the sake of consistency.[163]

For a record to be included in the modeling sample, the following conditions had to be met:

- It had to have valid SSA race/ethnicity data.
- It had to have a Census block location.
- The combination of coverages on the policy had to be "plausible," meaning the policy had to have one of the following combinations:
    - All four main coverages, or
    - Liability coverages and comprehensive, or
    - Liability coverages only.

---

[163] All of the analyses that did not require race or ethnicity data were also run on the full sample, and all results were very similar. These results are discussed in Appendix F.

- For the ChoicePoint score and FICO score risk models, the sample was limited to policies with a score. This was done because there were very few policies with a "no hit" or "thin file" that had SSA race and ethnicity data.[164]

In addition to the overall analysis sub-sample restrictions, there were additional restrictions for the individual coverage risk models:

- The earned car years for each record for the coverage being modeled had to be greater than zero and not greater than one.
- Total claims count had to be less than six. (This eliminated only a handful of records).

Table A.2 shows summary statistics for the database we analyzed. Column (a) shows statistics for the full sample of 1.4 million policies and 2.3 million vehicles. Column (b) shows the characteristics of the modeling sub-sample, and column (c) shows the characteristics of the modeling sub-sample when weights were applied to make the sub-sample nationally representative by geography, race, ethnicity, and income.

Comparing columns (a) and (b) shows that the sub-sample used for most of the analysis did not differ in any dramatic way from the full sample. This similarity is reassuring, especially given that some of the steps that produced the sub-sample could be quite non-random; in particular, the process of locating Social Security Numbers at Experian, which eliminated roughly ¼ of the original sub-sample of 400,000.

Comparing columns (b) and (c) shows that applying the nationally representative weights did affect some of the characteristics of the sub-sample. In particular, the share of people with missing values for many of the characteristics was quite different once nationally representative weights are applied. The likelihood that a characteristic is

---

[164] We did a separate analysis of "no hits" and "thin files" using Census race/ethnicity data. Those results are presented in part V of the report.

"missing" is determined by the information that the data providing firms collected and maintained. So, we reasoned that the change in the shares of many of the characteristics with unknown values reflected an effect of the nationally representative weights on the relative mix of the companies in the sample. As noted in Appendix F, all of the analyses were also run without the nationally representative weights, and this had very little effect on any of the results.

D.2 The Risk Models

The statistical models that the FTC constructed and used throughout the report are forms of Generalized Linear Models. These are fairly standard modeling techniques in the insurance industry. This section describes those techniques, and the specifics of how they were used to analyze the FTC database.

To better understand insurance claims risk, it helps to think of that risk as being made up of two components. The first component of risk is the probability that someone will file a claim. This is usually called "frequency." The second component of risk is the size of a claim, usually called "severity." Any risk factor, such as driver experience, geography, or credit history, could be correlated with either or both components of risk.[165] Because claims are generated in this way, claims data have certain distinct features. The data consist of a mix of a large number of zeros (policies with no claims) and a smaller number of positive dollar amounts. The mass of claims is centered around a relatively low number – the hundreds or low thousands of dollars – but claims can range into the tens or even hundreds of thousands of dollars. Both of these features – the many zeros and the long "tail" in the distribution of claims size – require the use of specialized statistical techniques.

---

[165] Some factors might affect both types of risk in the same direction. For example, someone who drives especially fast might be more likely to get into an accident, and any accident would probably be more severe than average. Other factors might affect the two types of risk in off-setting ways. For instance, someone with a very expensive car might be especially cautious and unlikely to have an accident, but face very high repair costs in the event of an accident.

There are two approaches to modeling risk. Either frequency and severity can be modeled separately, and the results combined, or total claims cost can be modeled in one step. Most of the analysis in the body of the report is concerned only with the total effects on risk of given variables, such as credit based insurance scores, so most of the analysis is done with total claims estimated in a single step. In the discussion of the predictive power of scores, however, separate results for frequency and severity are presented, as this may provide insights into how scores are predictive of risk. Whether risk is modeled in a single step or the two components of risk are modeled separately, the standard approaches are all built around Generalized Linear Models (GLMs).

D.2.1 Generalized Linear Models[166]

"Generalized Linear Models" are, as the name suggests, a class of statistical models that are generalized forms of standard linear models. GLMs generalize from linear models by allowing for the dependent variable to be distributed according to any member of the exponential family of distributions. GLMs also allow for the variance of the error term to vary with the mean of the distribution. Finally, GLMs allow the effects of explanatory variables to be a transformation of a linear function. The transformation is referred to as the "link function." A specific GLM model is defined by the link function and by the assumption made about the distribution of the dependent variable. The standard Ordinary Least Squares regression model is a special case of the GLM, with an identity link function and normally distributed errors.

---

[166] An excellent source for GLMs, especially in the context of modeling insurance claims, is Duncan Anderson, *et al*., A Practitioner's Guide To Generalized Linear Models, *CAS 2004 Discussion Paper* 1 – 116 (May 2005) (presented at Program CAS – Arlington).

D.2.1.1 Modeling "Frequency"

The standard approach to modeling frequency is called "Poisson regression," because it is a Generalized Linear Model (GLM) that uses the Poisson distribution. The Poisson distribution gives the likelihood that a certain number of events will occur in a given period of time, such as how many claims will be filed on an insurance policy during a year. The link function we used in our Poisson regression models was the natural log, so the regressions provided estimates of the multiplicative effects of the variables on risk. That is, the estimates show the effects of variables on relative risk, so an estimated effect of "2" means "predicted claims double when the variable takes this value."[167]

To implement a Poisson regression with the FTC database for a given coverage, the dependent variable was the number of claims for that coverage divided by the earned car years of that coverage.[168] To limit the effects of outliers, we dropped records that had more than six claims on a given coverage in a year.[169] Earned car years were also used as weights, because records with higher earned car years (that is, records that cover longer periods of time) contain more information about risk. Other weights were the sampling weights (which are necessary because the modeling sample was a stratified sub-sample of the original sample), the nationally-representative weights,[170] and, where necessary, the weights used to implement the race/ethnicity imputation. The explanatory variables in the model are listed below.

---

[167] The value "2" here would be the exponentiated coefficient estimate from the regression.

[168] Records with a positive count for a coverage but zero dollars paid out on claims for that coverage had the count set to zero. It is fairly common for a customer to file a claim that never results in a payment. For records with multiple claims and positive dollars paid on claims on a coverage, we cannot determine whether all of the claims resulted in payments, as we have only one variable on the total dollars paid on claims for each coverage. So, those records may overstate the number of claims that resulted in payments.

[169] This restriction caused very few records to be dropped, and it did not affect the results. Appendix F includes a discussion of the treatment of outliers.

D.2.1.2 Modeling "Severity"

The standard approach to modeling the severity of claims is to use a GLM with a Gamma distribution. The Gamma distribution is used because it has the features of the observed distribution of claims, all positive values with a relatively low central mass and a long tail of larger values. As with the Poisson regressions, we used a natural log link function for the Gamma GLM models, so the estimated effects from the model are multiplicative.

To implement a Gamma GLM for a given coverage in the FTC database, the sample was first limited to those records with claims on that coverage that resulted in payouts. The dependent variable for the severity regression was dollars paid out on claims for that coverage divided by the claim count for that coverage.[171] The size of claims was capped at the 99th percentile to mitigate the effects of outliers. The weights were the claim count, the sampling weight, the nationally representative weight, and the race-imputation weight, where needed. The explanatory variables were the same as for the frequency models, and are described below.

D.2.1.2 Modeling Total Claims Cost ("Pure Premiums")

---

[170] All of the analyses were also run without the nationally representative weights. As described in Appendix F, this had very little effect on any of the results.

[171] This may be affected by the problem of people with multiple claims on a single coverage, where we could not determine if all of the claims resulted in payments. For records with multiple claims and positive dollars paid on claims on a coverage, we cannot determine whether all of the claims resulted in payments, as we have only one variable on the total dollars paid on claims on a coverage. So, those records may overstate the number of claims that resulted in payments, which, in turn, will understate the average claim size for that record.

When frequency and severity are modeled separately, the results of the two models can be combined and the overall effect of a particular factor on expected dollars of claims can be calculated. It is also possible to model claims risk in a single step. This can be done by using a GLM with a "Tweedie" distribution.[172] The Tweedie distribution is a compound distribution of the Poisson and Gamma.[173] In essence, the Tweedie GLM approach addresses both the frequency effect and severity effect of risk factors in a single model. That is, it estimates the effect of a given factor on the total dollars of claims paid out per year of coverage.

To implement the Tweedie GLM with the FTC database for a given coverage, the dependent variable is dollars paid out on that coverage divided by earned car years. The same restrictions were placed on the dependent variable to limit outliers as were used in the frequency and severity models.[174] The weights in the pure premiums regressions were the same as for the frequency model: earned car years, the sampling weight, the nationally representative weight, and, where necessary, the race imputation weight. The explanatory variables were the same as those used in the frequency and severity models.

D.2.2 Bootstrapping Significance Tests

In several places in the analysis, we report the results of statistical significance tests. In each case, these tests were done using an approach known as "bootstrapping."[175] A bootstrap works by repeatedly drawing random samples, with replacement, from the analysis sample that are the same size as the analysis sample. Because these "pseudo-samples" are drawn with replacement, a record in the analysis sample may appear repeatedly, or not at all, in a given pseudo-sample. The parameter of interest is estimated for each pseudo-sample, and this is repeated many times. The confidence interval for the parameter can then be estimated simply by measuring the observed distribution of parameter estimates from all of the pseudo-samples.

---

[172] The distribution is named for M. C. K. Tweedie, who first introduced it in 1984. Tweedie MCK (1984). "An index which distinguishes between some important exponential families." In 'Statistics Applications and New Directions', Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. (Ed. JK Ghosh and J Roy) pp. 579-604. (Indian Statistical Institute: Calcutta).

[173] Estimating the Tweedie GLM models required choosing the value of a parameter of the distribution, P, that relates to the shape of the distribution and can vary between one and two. A standard approach is to use P=1.5, and that was used to produce the results presented in the report. We also tested values of P across the range from one to two, and the results of the models were not affected in any meaningful way.

[174] Using the same restrictions to avoid outliers did not eliminate all outliers from the pure premium models. Even though claim size and the nationally representative weights were capped, several claims became outliers when claim size, earned car years, and nationally geographic weights were combined. These were not excluded from the results reported in the body of the report. As discussed in Appendix F removing those records had no qualitative effects on the results, with one minor exception.

For example, bootstrapping was used to determine whether including race, ethnicity, and income controls had a statistically significant impact on the estimated risk impact of each score decile. This was done by first generating 500 pseudo-samples by drawing samples, with replacement, from the modeling sample.[176] The pseudo-samples were drawn at the policy level, so that any correlation in the unobserved risk across cars on the same policy would be accounted for in the bootstrapped confidence intervals. Once the pseudo-samples were generated, the risk models were estimated for each pseudo-sample, with and without controls for race, ethnicity, and income. The difference between the estimated risk for each score decile for the models with and without the controls was computed. Those differences are collected, and form the estimated distribution of the difference for each score decile. The 95% confidence interval for the difference for a given score decile can then be determined simply by measuring the value of the 2.5 percentile and 97.5 percentile of that distribution of estimated differences.

---

[175] A standard reference for the bootstrap is B. Efron and R. Tibshihanit, *An Introduction to the Bootstrap*, 1993, Chapman and Hall, Monographs on Statistics and Applied Probability (1993), at 57.

[176] The number of pseudo-samples is arbitrary. We found that our confidence intervals converged after 200 to 300 replications.

We also calculated robust standard errors for the parameter estimates of the GLM Tweedie models that took account of the fact that many records come from the same policies (*i.e.*,“clustering”). The resulting standard errors for the parameters of the models were very similar to those produced by the bootstrapping procedure. We rely on the bootstrap procedure, however, because statistical significance tests on the parameter estimates across different models cannot be done using the standard errors from those models (*e.g.*, comparing score decile parameter estimates across models with and without controls for race, ethnicity, and income).

D.2.3 Variables Used in the Risk Models

The following variables were used in the risk models. All variables were included in all models, except where otherwise indicated. All variables entered the models as indicator (“dummy”) variables. A number of variables have “missing” as one of the categories, and this category was included in the models. Whether a variable was missing for a record was determined by whether the company that provided that record had collected and maintained the information, and therefore when multiple records are missing the same variable it may mean they came from the same company. This complicates the interpretation of some variables, but may have the benefit of acting like an indicator variable for a particular company.

Credit-Based Insurance Score Decile

The credit-based insurance score decile of the score on the policy. Deciles were determined using property damage liability coverage earned car years as a weight, so each decile contains 10% of the property damage liability earned car years. The nationally representative weights were used when the score deciles were determined, and the same decile cut-points were used throughout the analysis.

Race/Ethnicity

Race and ethnicity category – from the SSA data, census data, and Hispanic surname match. As discussed above, this is a simple indicator variable for people for whom we have a post-1981 SSA race/ethnicity response, or who responded “Black” pre-1981. Individuals for whom we had only a pre-1981 response which was either “White” or “Other,” have separate records for each race or ethnicity that had a positive estimated probability, with a weight equal to the estimated probability. Race or ethnicity was included in models only where indicated.

Tract-Level Income

The median tract income relative to the Metropolitan Statistical Area median income. This variable takes the values of less than 80% of the MSA median (“low income”), 80% and greater but less than 120% of the MSA median (“middle income”), and 120% of the MSA median and greater (“high income”). Income was only used where indicated.

Age / Gender / Marital Status

The effects of age, gender, and marital status are all inter-dependent. The effect of age on risk varies with gender and marital status, the effect of gender on risk varies with age and marital status, etc. Fully interacting the three variables, however, leads to literally hundreds of possible combinations. To reduce the set of controls used in the models, we created groupings of age/gender/marital status that were of similar risk. We first created a set of seven age ranges. The age ranges were determined by estimated frequency risk models with varying age bands, which in turn were based in part on the public rate filings of several firms. The chosen categories were interacted with gender and marital status. Because gender could take three values (male, female, unknown) and marital status could take four values (single, married, divorced or widowed, unknown), this produced a total of 7x3x4=84 cells. We then ran risk models for each of the four major coverages using all 84 cells. The results showed that the effects on risk of the age/gender/marital status categories were fairly similar across the accident-related coverages (the liability coverages and collision), but somewhat different for comprehensive coverage. We therefore created two sets of age/gender/marital status categories. After examining the estimated risk effects of the 84 cells, we created nine risk categories for the accident-related coverages and six categories for comprehensive. This was done based on the predicted risk for the 84 cells, with attention paid to creating "reasonable" categories made up of cells that were "close" to each other on a grid of age/gender/marital status.

Territorial Risk

The territorial risk variable was calculated by EPIC using three-year average property damage liability claims for the five companies, by ZIP-code. This is described in more detail in Appendix C. Territorial risk entered the model as quintiles, five groups that each contain 20% of the vehicles in the sample, weighted by property damage liability coverage earned car years. As described in Appendix F, using deciles instead of quintiles did not affect the risk models.

CLUE Data

The CLUE data contains information on the number and size of claims for the full range of coverages. Several variables were used to capture that information for inclusion in the risk models.[177]

CLUE Data – Prior Uninsured Motorist / Underinsured Motorist Claims

The number of claims that involved an uninsured or underinsured motorist claim with a positive dollar value in the prior three years. It takes the values of "0" and "1 or more."

CLUE Data – Prior Bodily Injury / Property Damage Claims

The number of claims that involved a bodily injury or property damage claim with a positive dollar value and did not involve uninsured or underinsured motorist claims, in the prior three years. This variable takes the values "0," "1," "2," and "3 or more."

CLUE Data – Prior Collision / Medical Payments / Personal Injury Claims

The number of claims that had a collision, medical payments or personal injury claim with a positive dollar value, and did not have uninsured or underinsured motorist, bodily injury, or property damage claims, in the prior three years. This variable takes the values "0," "1," and "2 or more."

CLUE Data – Prior Comprehensive-Only Claims

Number of claims involving only comprehensive coverage with a positive dollar value, in the three prior years. This variable takes the values "0," "1," "2," and "3 or more."

CLUE Data – Prior Towing and Labor-Only Claims

Number of claims involving only towing and labor with a positive dollar value, in the prior three years. This variable takes the values “0,” “1,” “2,” and “3 or more.”

---

[177] Because we received prior-claims data only for people who had the same address in the company data and in the CLUE data, the prior claims data used in the FTC's analysis may be more limited than that used by companies when they underwrite and rate policies. Companies can ask applicants for prior addresses, and submit those addresses to be matched, as well.

CLUE Data – Prior Rental Reimbursement Claims

Number of claims involving rental reimbursement, in the prior three years. This variable takes the values “0,” “1,” and “2 or more.”

Number of Accidents

Number of accidents indicates the number of “chargeable” accidents that occurred prior to the beginning of the policy period.[178] This variable came from the companies, and may only reflect claims made policies at that company. The variable is missing for a large portion of the sample. The definition of “chargeable accident” may vary by company and by state, but is usually based on a dollar threshold and often on whether the driver was found to be at fault. For an accident to be considered chargeable, it must typically have occurred in the previous three years. This variable takes the values of “zero,” “one or more,” and “unknown.”

Number of Violations

The number of violations indicates the sum of major and minor moving violations for the driver assigned to a car that occurred prior to the beginning of the policy period. The definition of major violation may vary by company and by state. Typically, this variable only includes major and minor violations in the past three years. This variable takes the values “zero,” “one or more,” and “unknown.”

Tenure

Tenure is the number of years the customer had been with the company. Each year of tenure is a separate category for years 1 through 14, and then years of tenure are combined into categories for “15 or 16 years,” “17, 18, or 19 years,” and a final category for tenures of 20 years or more.

---

[178] This variable may in many or all cases exclude accidents that occurred while the consumer was a customer of a different firm, which is one reason the CLUE data provides important additional information.

Property Damage Liability Limits

This is the maximum amount customers would be reimbursed for a property damage liability claim. It was used only in liability regressions. Property Damage liability limits takes the values “$5,000 - $10,000,” “$15,000 - $20,000,” “$25,000 - $45,000,” “$50,000 - $80,000,” “$100,000 - $200,000,” “$250,000 - $325,000,” “$500,000 - $2,000,000,” and “missing or zero.” (Note that when these ranges are non-contiguous there were no policies in the database with values between the ranges.)

Bodily Injury Liability Limits

Bodily injury limit is the maximum amount customers would be reimbursed for bodily injury claims. There are two limits on bodily injury liability, the per-person limit and the total cost limit per occurrence. The two limits are highly correlated, so we based our bodily injury liability limit variable on the per-person limit. It was used only in liability regressions. It takes the values “$10,000,” “$12,500 - $15,000,” “$20,000,” “$25,000 - $40,000,” “$50,000 - $75,000,” “$100,000 - $150,000,” “$200,000 - $250,000,” “$300,000 - $400,000,” “$500,000 - $2,000,000,” and “missing or zero.” (Note that when these ranges are non-contiguous there were no policies in the database with values between the ranges.)

State Minimums

State minimums indicated whether the policy had only the minimum liability coverage required by law. It takes the values of “yes” and “no”. The FTC created this variable by comparing the liability limit variables with data on state legal minimum liability requirements. Information on state minimums, as of 2000, came from the NAIC 2001/2002 Auto Insurance Database Report.[179]

Collision Deductible

This is the deductible for collision claims. It takes the values "$0 - $50," "$100 - $150," "$200," "$250 - $400," "$500," "$1,000 - $1,500," and "missing." This variable was used only in collision regressions. (Note that when these ranges are non-contiguous there were no policies in the database with values between the ranges.)

Comprehensive Deductible

This is the deductible for comprehensive claims. It takes the values "$0 - $25," "$50," "$100 - $150," "$200," "$250 - $300," "$400 - $750," "$1,000 - $5,000," and "missing." This variable was used only in comprehensive coverage regressions. (Note that when these ranges are non-contiguous there were no policies in the database with values between the ranges.)

Annual Mileage

Estimated annual mileage as reported by the customer. It takes the values of "7,500 miles or less," "more than 7,500 miles," and "unknown."

Principal / Occasional Driver

Principal or occasional operator identifies whether the driver assigned to a vehicle was the primary user of the vehicle, or only used it occasionally. It is an indicator that is typically used only for young drivers. The variable had categories for "principal (driver)," "occasional (driver)," and "unknown".

---

[179] National Association of Insurance Commissioners, "Auto Insurance Database Report 2001/2002" (2004).

Use

Vehicle usage reflects whether the vehicle was used primarily for “pleasure,” “farm,” “business,” “travel to work,” “all other uses,” or whether the use was “unknown.”

Homeowner

Indicates whether the customer owned a home. It takes the values “yes” and “no.”

Multi-line Discount

Multi-line discount designates whether a customer had multiple types of insurance with their auto carrier. The discount is commonly applied when a customer purchases homeowners insurance from the same company. Multi-line discount takes the values of “yes,” “no,” and “unknown.”

Multi-Car

Multi-car indicates whether there were multiple cars in the household covered by the same insurer. It takes the values “yes,” “no,” and “unknown.”

State

State where the vehicle was principally garaged.

Model Year

Model year of the vehicle. Each model year is a separate category, except the following groups of years: “2001 – 2002,” “1981 – 1984,” and “1980 or older.”

Body Type

Data from Edmund’s on the vehicle type. Body type takes the values “convertible,” “coupe,” “extended or crew cab pickup,” “regular cab pickup,” “four-door SUV,” “two-door SUV,” “hatchback,” “passenger minivan,” “wagon,” “sedan,” and "unknown."

Restraint System

Data from Edmund's on airbags and seat belts. Restraint system takes the values "only passive seatbelts," "only active seatbelts," "seatbelts and driver's front airbag," "seatbelts and driver and passenger front airbags," "more than seatbelts and front airbags," and "unknown."

Displacement

Data from Edmund's on the size of the engine in the vehicle. Engine displacement is an indicator of the power of the engine. It takes the values "less than 2.7 liters," "2.7 – 4.3 liters," "More than 4.3 liters," and "unknown."

APPENDIX E

THE SCORE BUILDING PROCEDURES

APPENDIX E. Score Building Procedures

E. 1. Developing the FTC Base Model

The FTC credit-based insurance score-building methodology produces "pure premium" scoring models. That is, the models are developed to predict total dollars paid out on claims on a policy in a year.[180] To have a single scoring model that predicts losses for any of the four major coverages, we combined total claims across coverages into a single measure of losses.[181]

The steps for building a credit-based insurance scoring model are first described, and then the logic underlying the procedure is discussed.

- An ordinary least squares model ("OLS model") is run using total dollars of claims as the dependent variable, and the 180 credit history variables as the explanatory variables.[182] The results of the OLS model are used to generate a "proto-score."

- A Tweedie GLM model is run, using total dollars of claims as the dependent variable, and all the standard risk variables and the proto-score as the explanatory variables.[183] Predicted total dollars of claims are calculated for each record using the results of the Tweedie GLM model.[184] An "adjusted claims" variable is calculated by dividing actual total dollars of claims by predicted total dollars of claims.

- Each credit history variable is then divided into optimal "bins." This is done using an approach developed by staff of the FRB. The relationship between each credit history variable and adjusted claims is evaluated separately. First, the credit history variable is divided into the two categories that create the biggest difference in mean adjusted claims between the two categories. These categories are then divided into additional categories, until the point where further divisions would not lead to statistically significant differences in mean adjusted claims across new categories.[185]

---

[180] Because many of the records are for less than a full year, total dollars of claims are adjusted for the period of time each car was actually covered by one of the companies in the sample.

[181] Claims on first party medical coverages – MedPay and personal injury protection – are also included in the "total losses" variable.

[182] The credit history variables were first converted from continuous variables into discrete variables. This was done using a simple rule of thumb of dividing the values into "bins" that each contains at least roughly 10% of the sample. (So, if 50% of the sample had a value of zero for a given variable, there would be one category for "zero," and up to five additional bins.)

[183] Because we are combining claims from across coverages, we also include dummy variables indicating whether the policy included collision, comprehensive, MedPay, and/or personal injury protection coverage.

[184] The "proto-score" is used in the model estimation as a control, but is not used when the predicted pure premium is calculated. The use of a "proto-score" in this way follows a suggestion from several score builders at firms. It is done simply to minimize the effect of other variables that are correlated with score, such as age, picking up variation that would be attributed to score if score were included in the model.

- A forward-selection OLS model is run, with adjusted claims as the dependent variable, and the binned credit history variables as the candidate explanatory variables. The process works by first choosing the variable that, on its own, is most predictive of risk, based on an F-test. The next variable chosen is the variable that adds the most predictive value when used in a model with the first variable chosen (again, based on an F-test). This process continues, with credit history variables being added, one by one, until a pre-determined threshold is reached.

- A Tweedie GLM model is run with actual total dollars of claims as the dependent variable, and the standard risk variables and the "winning" credit history variables as the explanatory variables.

- The coefficients on the credit history variables from the Tweedie GLM estimated in the previous step are used to generate a scorecard for the "FTC credit-based insurance scoring model."

The underlying logic of this procedure is that we are attempting to find the set of credit history variables that best predict total dollars of claims, after controlling for non-credit risk variables. The non-credit risk variables are initially included in the model by adjusting total dollars of claims by a measure of risk based on these variables. Steps one and two do this. The third step, the binning of the credit-history variables, is done for two reasons. (The alternative would be to keep the credit-history variables as continuous variables.) Dividing the values into bins is a simple way of allowing the effects of the variables to vary in complex non-linear ways over the range of values. Using bins also makes the scorecard – the tool for actually calculating a score – much simpler than would other ways of allowing non-linear effects.

---

[185] Two restrictions are placed on the binning process. The first is that no bin could be less than ½% of the total sample. This is done to avoid "over fitting" the data, and to avoid convergence problems when binned data are used in the Tweedie GLM stage. The binning procedure was also run using either a monotonicity requirement, meaning that average claims must either rise or fall across the range of bins, or a "single-turning" requirement, meaning that if average adjusted claims were not monotonic, they could first go up and then down, or vice-versa, but not go up-down-up or down-up-down, etc. Both restrictions led to the same set of optimal bins.

The fourth and fifth steps are the core of the score-building process. First, the most predictive credit-history variables are determined by the forward-selection procedure. The forward-selection procedure runs a separate OLS model regression, with adjusted claims as the dependent variable, for every credit-history variable (*i.e.*, 180 separate regressions). It then determines which credit-history variable provides the most predictive power. It then runs through that same process, and chooses the variable that adds the most predictive power to a model that includes the "winning" variable from the first step. This process continues, adding variables one-by-one, until it hits some stopping rule.[186] We used two stopping rules. The first was that if the estimated effect on adjusted claims of the next potential variable was not statistically significantly different from zero ("no effect") at the 10% level, the procedure stopped. This approach tended to produce a model with a very small number of variables, fewer than ten. We also used an alternative where the procedure continued until it had selected the first fifteen "winning" variables. Fifteen was chosen arbitrarily, based on scorecards we reviewed and discussions with professional score builders and staff at the Federal Reserve Board.

The final step in the score building process is calculating the scorecard. This is done by estimating a Tweedie GLM with actual total claims, instead of adjusted total claims, as the dependent variable. All of the non-credit risk variables are included in the model, along with the "winning" credit history variables. The scorecard is made up of the estimated coefficients on the credit history variables. The scorecards we report show the inverse of the exponentiated coefficients. A score is calculated by multiplying together the coefficients for each credit history variable, and this produces the inverse of the predicted relative risk. The coefficients must be exponentiated because the Tweedie GLM has a log-linear functional form. We use the inverse of the coefficients so that a higher score will be associated with a lower predicted risk.

---

[186] Ideally, the forward selection procedure would be run using a Tweedie GLM model, as that is the preferred way of modeling total dollars of insurance claims. Maximum likelihood procedures are apt to "crash," however, especially when run on data with many highly-correlated variables, like credit history data. (It is common in the industry to use some form of OLS-based variable selection procedure.) Our approach is a compromise. We use the OLS model for the forward-selection procedure, which determines the "winning" variables, but estimate the final scoring model using a Tweedie GLM model of the actual pure premiums with all of the non-credit variables.

E.2 Developing "Race Neutral" Models

The FTC used two approaches to controlling for race, ethnicity, and income in the score-building process. One approach was to include controls for race, ethnicity, and income in the forward-selection step, when the "winning" credit history variables were chosen.[187] This means that the variables were not chosen because of a correlation with race, ethnicity, or income. Race, ethnicity, and income controls were also included when the final Tweedie GLM was run to generate the scorecard. So, any relationship between risk and race, ethnicity, and income was controlled for, and would not be picked up by the weights on the credit history variables. (Note that while race, ethnicity, and income are included in the model that determines the scorecard, they are not themselves used to calculate a score.)

---

[187] An alternative approach we used was to include race, ethnicity, and income controls in the step of the model-building process when the "adjusted pure premium" is calculated. The adjusted pure premium was therefore adjusted for those variables. The binning of the credit history variables was therefore done in a way that was purged of any relationship between race, ethnicity, and income, and claims. In addition, the forward-selection process was done with the race adjusted pure premium as the dependent variable. So, the credit history variables were chosen for the model using a dependent variable that was adjusted for race, ethnicity, and income. This approach gave very similar results to results of the model discussed in the body of the report.

The other approach was to build the model using only non-Hispanic whites. This was done by limiting the development sample to people who answered "White, Non-Hispanic" to the post-1981 SSA questionnaire and the records that represent the "non-Hispanic white" imputed probabilities of people for whom we only have pre-1981 SSA data (which include the weight from the imputation process).

E.3 Discounting Variables for Differences across Racial and Ethnic Groups

To force the model-building procedure to produce models with smaller differences across racial and ethnic groups, we modified the forward selection step to take those differences into account. Normally, the forward selection step runs a series of OLS regression models, with adjusted total claims as the dependent variable and credit history variables as the explanatory variables. One regression is run for each credit history variable. The credit history variable with the largest impact on predicted risk at each step, as measured by an F-test, is added to the set of "winning" variables.

This step was modified by also running an OLS regression for each credit history variable with race and ethnicity as the dependent variable. Race and ethnicity was captured using indicator variable for whether the individual was non-Hispanic white or minority (*i.e.*, all minority groups were combined into one category, to simplify the modeling). The R^2 statistics were then calculated for the risk OLS model and the race/ethnicity OLS model, and used jointly to choose winning variables. The R^2 statistic from the risk equation is a measure of how much power the credit history variable has to predict risk. The R^2 statistic from the race and ethnicity model is a measure of how much the credit history variable differs by race and ethnicity. We used these two measures to choose variables for the model in a variety of ways. The approach described in the body of the report was to first normalize the R^2 statistics within each set of regressions – the risk regressions or the race and ethnicity regressions – by dividing the R^2 for the regression for each variable by the largest R^2 in that set of regressions. That is, the R^2 statistics from the risk regressions for each credit history variable were divided by the largest R^2 from all of the risk regressions, and similarly for the race and ethnicity regressions. We then compared the normalized R^2 statistics to select the variables to include in the model.[188]

---

[188] The model described in the body of the report was determined by subtracting twice the normalized R^2 of the race and ethnicity regression for each variable from the normalized R^2 of the risk regression for that variable. At each step, we chose the variable with the largest difference as the winning variable. Taking the difference between the normalized R^2 statistics, without doubling the normalized R^2 from the race/ethnicity regression, resulted in a model with much larger differences across racial and ethnic groups. Using the ratio of the R^2 statistics from the two regressions resulted in a model that was very similar to that discussed in the body of the report.

APPENDIX F

ROBUSTNESS CHECKS AND LIMITATIONS OF THE ANALYSIS

APPENDIX F. Robustness Checks and Limitations of the Analysis

The FTC conducted numerous additional analyses to confirm the results presented in the body of the report, and to test whether those results are robust to the credit score used, the database used, the use of a sub-sample, and a variety of modeling decisions. There remain several limitations of the database and the analysis that could not be fully addressed through these robustness checks.

F.1 Limitations of the Data and the Analysis

No Information on People who did not Obtain Insurance

The FTC did not have information on insurance applicants who were denied coverage by the firms that provided data. We could therefore not directly evaluate the impacts of credit-based insurance scores on consumers' ability to obtain insurance from a given firm. However, the analysis of state residual markets data in NAIC reports shows that scores do not appear to have adversely affected consumers' ability to obtain insurance through the normal, voluntary market for automobile insurance.

Single National Model

Underwriting and rating plans are determined by firms below the national level, and often at the state level. The FTC's analysis includes controls for state, but does not separately model risk by state. The results of our national model may differ from the results of separate state-level models, especially if the effects of particular risk variables differ across states.

Pooled Company Data

The FTC risk models were estimated using pooled data from multiple firms.

Individual firms estimate the risk posed by their customers, and the results of models estimated using data from a single company may differ from those of a model estimated using pooled data.

Sub-Sample of Industry

The FTC database includes data from five firms that together represented over ¼ of the entire automobile insurance market as of 2000. Despite having data from a fairly large share of the market, we know that this sample likely under-represents the highest-risk segment of the market. (An analysis that focuses on a sub-sample of the riskiest policies in our database is presented in section F.2, below.) In addition, there may be other ways in which these firms differ from the market, as a whole.

Territorial Risk Variable

The territorial risk variable in the FTC database is based on ZIP-code average property damage liability claims. It is a powerful predictor of risk for property damage liability, bodily injury liability, and collision coverages, but it may differ from the territorial risk measures used by individual firms. More importantly, this territorial risk variable is not a powerful predictor of risk for comprehensive coverage. As discussed in the text, this is likely to lead to over-estimates of the relationship of both score and demographic characteristics like race, ethnicity, and neighborhood income to comprehensive coverage risk.

F.2 Robustness Checks

FICO Score

The credit-based insurance score results reported in the body of the report are for the ChoicePoint Attract Standard Auto score. All of the analyses were also run using the

FICO "Standard Auto, Greater than Minimum Limits" credit-based insurance score. The results were similar, both qualitatively and quantitatively, to the results for the ChoicePoint score.

No Nationally Representative Weights

The level of racial, ethnic, and income diversity of the sample could affect the results of the "proxy" analysis. The analysis in the body of the report was done using a sample weighted to match the racial, ethnic, and neighborhood income distribution of the national population of car owners. While this seems a reasonable approach, that population may have a different racial, ethnic, or income mix than the national population of car insurance customers, or the mix of the pool of customers of any individual firm. We also did the analysis without using the tract and race weights that make the sample nationally representative. The results were qualitatively very similar to the results from using the weights. The impact of scores on the estimated risk of African Americans and Hispanics was slightly larger, with the impact on African Americans being an average increase of 11.6% (versus 10.0% with weights) and for Hispanics 5.8% (versus 4.2% with weights). The estimated proxy effect was very similar.

Outliers

We suspected that policies with more than six claims on a coverage may have reflected data errors, so those policies were dropped from the analysis reported in the body of the report. Leaving those policies in did not affect the results of the analysis.

The use of nationally representative weights resulted in several claims becoming outliers, despite the capping of those weights at the 99.95th percentile. There were four people with large claims and small earned car years who lived on Census tracts that were highly under-represented in the database whose claims became outliers when the Census tract weights were applied. Two of these had no impact on any results. These were a collision claim paid to a Hispanic consumer in the lowest score decile, and a comprehensive claim paid to an African American consumer in the 3rd-lowest score decile. There were two outliers that did have a small impact on the results described in the body of the report. There was one bodily injury liability claim, filed by a non-Hispanic white consumer in the second score decile (the second from the bottom) that became an outlier. Capping the weighted value of the claim at the size of the next-largest weighted claim reduced the estimated risk effect of the second decile in the bodily injury liability model by several percentage points. This did not affect any other results of the analysis. There was one comprehensive claim, filed by an African American consumer in the 9th score decile (second from the top) that became an outlier when the nationally representative weight was applied. This did not affect the estimated risk effect for the 9th decile in the overall comprehensive claims model, and therefore does not affect any of the overall results of the analysis. It did have a large effect on the estimated risk effect of the 9th decile for African Americans when race and ethnicity were interacted with score deciles (this is shown in Figure 14). Capping this claim brought that estimated risk effect down somewhat, but only when the observation was dropped did the estimate fall in line with the surrounding deciles. In any case, the estimated risk for the 9th decile for African Americans was not statistically significantly different from that of the overall sample, even when the outlier was not capped.

Full-Sample Models

With the exception of the analysis of the CLUE claims data, the results in the body of the report are based on a sub-sample of records. Much of the analysis required the SSA race ethnicity data, and therefore could be done only with the sub-sample for which we obtained those data. We also estimated the basic risk models, without race/ethnicity/income controls, on the complete sample. The results were very similar to the results from the sub-sample that are described in the body of the report.

Census-Only Race and Ethnicity Data

In the body of the report, we combined data on race and ethnicity from three sources: Social Security Administration records, a Hispanic Surname match, and Census data. We also estimated models using only Census race and ethnicity data, measured at the Census block level. This resulted in a weaker relationship between race/ethnicity and claims risk, which, in turn, resulted in a smaller estimated "proxy effect." These results are what would be expected when race and ethnicity are measured less precisely.

Absolute Income Measure

The results presented in the body of the report that relate to income are based on assigning people to one of three income categories based on the median income of the Census tract where the person lived relative to the median income in their MSA. To determine whether using relative income instead of absolute income affects the results of our analysis, we re-ran the analysis using three categories based on tract median income, not relative to the MSA median. This did not affect the results of the analysis.

Race and Ethnicity Imputation Cut-Offs

As discussed in Appendix C, when multiple data sources were used to impute the race and ethnicity of people for whom we only had a pre-1981 SSA race/ethnicity answer, we imposed a minimum cut-off on the predicted likelihood that someone was of a given race or ethnicity. When the estimated likelihood of being of a particular race or ethnicity was very low, we set the probability to zero. To test whether this decision affected the results, we re-ran the analysis without using the cut-off. This did not affect the results.

High-Risk Sub-Group

Because of the way the sample was drawn by the companies, the FTC database probably under-represents the highest-risk portion of the automobile insurance market. In an attempt to determine whether our analysis would extend to that portion of the market, we estimated risk models limited to the riskiest people in our database, as determined by non-credit factors. To do this, we first ran a risk model without credit score, on the full model sample, that combined claims from the four major coverages. We then predicted each individual's expected total claims (their risk), and created a sub-sample consisting of the 20% of the sample with the highest predicted total dollars of claims. We then ran risk models for each of the four major coverages that included credit scores on the "risky" sub-sample. The estimated relationships between risk and score for the sub-sample were similar to the relationships in the overall sample.

Estimating Total Losses by Modeling Frequency and Severity Separately

Most of the results in the report are from Tweedie GLMs of total dollars of claims. In addition, we modeled total dollars of claims by separately modeling frequency of claims, using Poisson regressions, and severity of claims, using Gamma GLMs, and then combined the estimates from the two models. The estimated relationships between score and risk from combining the results from these two models were essentially identical to the results from the single-step model.

Single Combined-Coverage Model

In the body of the report, we present results from analyzing each type of automobile coverage separately. In addition to the separate models by coverage, we estimated a combined-risk model for the four major coverages. This was done by summing claims on the four major coverages into a single claims variable. Indicator variables were included to control for differences in the set of coverages purchased by consumers. Scores were predictive of risk in this combined-coverage model, and the effects of scores on the predicted risk of different racial and ethnic groups from the combined-risk model were very similar to the results from combining the results from the separate coverage models. The overall "proxy" results for scores were also similar to the results from combining the results from the separate coverage models.

"Tiering"

The risk models used in the body of the report are single-equation models, where all risk factors enter into the single equation. Some firms use credit-based insurance scores to determine the risk category in which a customer is placed. This may allow the effects of non-credit risk variables to vary depending on a person's score (essentially interacting score with other risk variables). To determine whether this would affect the results of our analysis, we divided the sample into three groups based on score. We then ran separate risk models for the three groups, with and without score, and measured the impact on predicted risk for different racial and ethnic groups. The results were very similar to the single-model approach used in the body of the report.

Number of Score Categories

We use score deciles throughout the report. To test whether the choice of deciles was important to the results, we re-ran the analysis using 20 categories of scores ("ventiles"). The results for predicted risk, predicted impacts on minorities, and the results relating to "proxy effects" from using ventiles was very similar to the results from using deciles.

Number of Geographic Risk Categories

In the results reported in the body of the report, we use controls for geographic risk that assign people to five categories ("quintiles"). To test whether the choice of quintiles was important to the results, we re-ran the analysis using ten categories of geographic risk (deciles). Using deciles of geographic risk, instead of quintiles, did not affect the results.

APPENDICES
TABLES

TABLE A1.
Development of Nationally Representative Weights: Share of Vehicles by Race, Ethnicity and Neighborhood Income

	Census	FTC Database		
	(a)	Unweighted (b)	With Tract Weights (c)	With Tract and Race Weights (d)
Race				
African Americans	8.4%	4.3%	6.0%	8.4%
Hispanics	7.8%	2.8%	3.7%	7.8%
Asians	3.1%	3.1%	2.9%	3.1%
Non-Hispanic Whites	80.8%	89.8%	87.5%	80.8%
Income				
Low	18.2%	12.3%	17.6%	19.2%
Medium	50.6%	44.0%	51.0%	50.3%
High	31.3%	43.7%	31.4%	30.5%

Notes:

1) Percentages are relative to the group of consumers included in these calculations.

2) The tract weights were calculated using the ratio of the share of vehicles in the 2000 Census in each tract divided by the share of vehicles in the FTC database in each tract. The subsequent race weights are simply the ratio of the share of each race group in the Census data over the share of each race group in the FTC database, after applying the tract weights. See Appendix C for details on the development of the weights.

3) The final proportions differ slightly from those reported in the table on the sub-sample used for model estimation and analysis because that sample has several additional minor restrictions that were not applied to the sample used to develop the weights.

TABLE A2.
Summary Statistics for the Full FTC Database and the Sub-Sample Used for Model Estimation and Analysis

	Full Sample (a)	Model Sub-Sample (b)	Model Sub-Sample With Nationally Represenative Weights (c)
	(Median or Percent)	(Median or Percent)	(Median or Percent)
Gender			
Male	29.8%	29.2%	25.8%
Female	31.4%	32.1%	28.9%
Unknown	38.8%	38.7%	45.3%
Marital Status			
Single	12.3%	13.1%	12.3%
Married	31.6%	33.1%	27.4%
Divorced / Widowed	2.4%	2.6%	2.9%
Unknown	53.7%	51.1%	57.5%
Accidents			
Zero	56.8%	59.7%	60.7%
One or More	4.5%	4.9%	4.7%
Unknown	38.6%	35.4%	34.6%
Miles Driven			
<7500	22.1%	22.0%	22.5%
>7500	50.4%	50.6%	55.0%
Unknown	27.6%	27.5%	22.5%
Multi-Line Discount			
Yes	34.5%	34.1%	36.6%
No	35.3%	34.8%	40.4%
Unknown	30.3%	31.1%	23.1%
Principal Operator			
Yes	28.2%	27.8%	27.0%
No	6.0%	5.7%	5.9%
Unknown	65.8%	66.5%	67.1%
Use			
Business	0.7%	0.6%	0.6%
Farm	0.8%	0.6%	1.0%
Pleasure	42.3%	43.1%	44.0%
Work	15.7%	16.9%	18.5%
All Other	12.8%	9.6%	11.2%
Unknown	27.8%	29.2%	24.8%
Homeowner			
Yes	55.6%	56.3%	52.5%
No	44.4%	43.7%	47.5%
Multiple Cars			
Yes	61.8%	60.1%	53.0%
No	14.5%	14.8%	14.9%
Unknown	23.7%	25.1%	32.0%

(continued. . .)

TABLE A2.
Summary Statistics for the Full FTC Database and the Sub-Sample Used for Model Estimation and Analysis (Continued)

	Full Sample (a)	Model Sub-Sample (b)	Model Sub-Sample With Nationally Represenative Weights (c)
	(Median or Percent)	(Median or Percent)	(Median or Percent)
Major Violations			
Positive	0.3%	0.3%	0.4%
Zero	64.6%	64.9%	59.5%
Unknown	35.1%	34.8%	40.1%
Minor Violations			
Positive	5.1%	5.5%	5.1%
Zero	54.3%	53.9%	47.2%
Unknown	40.6%	40.6%	47.8%
Vehicle Body Type			
Convertible	1.6%	1.6%	1.2%
Coupe	5.5%	5.8%	6.2%
Ext/Crew cab pickup	4.4%	4.4%	4.8%
Four-door SUV	9.7%	9.8%	8.5%
Hatchback	3.7%	3.8%	3.7%
Passenger MiniVan	5.5%	5.7%	5.5%
Regular Cab Pickup	3.6%	3.5%	4.5%
Two-door SUV	1.8%	1.8%	1.8%
Wagon	2.9%	3.0%	2.1%
Sedan	31.0%	31.5%	30.2%
Unknown	30.4%	29.0%	31.5%
Restraint System			
Driver's front airbags	10.7%	10.9%	10.8%
Driver/Psgr front airbags	36.3%	37.4%	35.5%
Just active belts	12.1%	12.0%	12.5%
Just passive belts	5.6%	5.7%	6.1%
More than front airbags	5.0%	5.0%	3.5%
Unknown	30.4%	29.0%	31.5%
Prior Claim[†]			
Under & Uninsured Motorist	1.6%	1.7%	2.0%
BI & PD	14.4%	15.1%	14.4%
Coll., Med., & PIP	12.9%	13.9%	13.5%
Comprehensive	19.3%	20.6%	19.8%
Towing and Labor	6.7%	7.2%	8.1%
Rental Reimbursement	7.3%	8.1%	8.4%
None of the above	60.9%	58.3%	58.9%
Age	47	46	46
Share Unknown	*12.6%*	*12.3%*	*11.7%*
Tenure	10	10	8
Share Unknown	*11.7%*	*11.3%*	*12.8%*

(continued. . .)

TABLE A2.
Summary Statistics for the Full FTC Database and the Sub-Sample Used for Model Estimation and Analysis (Continued)

	Full Sample (a)	Model Sub-Sample (b)	Model Sub-Sample With Nationally Represenative Weights (c)
	(Median or Percent)	(Median or Percent)	(Median or Percent)
Property Damage Liability Limit	$50,000	$50,000	$50,000
Share Unknown	*3.2%*	*3.1%*	*2.0%*
Bodily Injury Liability Limit	$100,000	$100,000	$100,000
Share Unknown	*3.6%*	*3.4%*	*2.3%*
Collision Deductible	$500	$500	$300
Share Unknown	*0.0%*	*0.0%*	*0.0%*
Comprehensive Deductible	$200	$200	$100
Share Unknown	*0.0%*	*0.0%*	*0.0%*
Model Year	1994	1994	1994
Share Unknown	*0.8%*	*0.4%*	*0.3%*
Coverage Combinations			
All Four Main Coverages	77.3%	82.6%	80.6%
Liability and Comprehensive	13.3%	13.3%	15.4%
Liability Only	4.1%	4.1%	4.1%
Other Coverage Combinations	5.4%	0.0%	0.0%
Race/Ethnicity			
African American	NA	4.3%	8.4%
Hispanic	NA	2.8%	7.7%
Asian	NA	3.1%	3.1%
Non-Hispanic White	NA	89.9%	80.8%
Number of Policies	1,434,041	275,509	275,509
Number of Vehicles	2,284,330	458,940	458,940
Total Car Years	1,808,584	399,100	399,100

†: Some Prior Claims categories are not mutually exclusive, therefore the shares can add up to more than 100%

Note: See Appendix C for details on the data sources and the construction of the database. See Appendix D for a discussion of how the sub-sample used for model estimation and analysis was chosen.